

DEX-0314

BH

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>7</sup> : <b>C12N 9/00, C07K 14/00</b>	<b>A3</b>	(11) International Publication Number: <b>WO 00/21986</b> (43) International Publication Date: 20 April 2000 (20.04.00)
<p>(21) International Application Number: PCT/US99/23315</p> <p>(22) International Filing Date: 6 October 1999 (06.10.99)</p> <p>(30) Priority Data: 09/169,289 9 October 1998 (09.10.98) US</p> <p>(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application US 09/169,289 (CIP) Filed on 9 October 1998 (09.10.98)</p> <p>(71) Applicant (for all designated States except US): INCYTE PHARMACEUTICALS, INC. [US/US]; 3174 Porter Drive, Palo Alto, CA 94304 (US).</p> <p>(72) Inventors; and (75) Inventors/Applicants (for US only): WALKER, Michael, G. [CA/US]; Unit 80, 1050 Borregas Avenue, Sunnyvale, CA 94089 (US). VOLKMUTH, Wayne [US/US]; 783 Roble Avenue, #1, Menlo Park, CA 94025 (US). KLINGLER, Tod, M. [US/US]; 28 Dover Court, San Carlos, CA 94070 (US).</p> <p>(74) Agents: BILLINGS, Lucy, J. et al.; Incyte Pharmaceuticals, Inc., 3174 Porter Drive, Palo Alto, CA 94304 (US).</p>	<p>(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).</p> <p><b>Published</b> <i>With international search report.</i></p> <p>(88) Date of publication of the international search report: 13 July 2000 (13.07.00)</p>	
<p>(54) Title: MATRIX-REMODELING GENES</p> <p>(57) Abstract</p> <p>The invention provides novel matrix-remodeling genes and polypeptides encoded by those genes. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating or preventing diseases associated with matrix remodeling.</p>		

*FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Larvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

# INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 99/23315

A. CLASSIFICATION OF SUBJECT MATTER  
IPC 7 C12N9/00 C07K14/00

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	MATRISIAN, L. M.: "METALLOPROTEINASES AND THEIR INHIBITORS IN MATRIX REMODELING" TRENDS IN GENETICS, vol. 6, no. 4, 1990, pages 121-125, XP000644304 * figure 2; page 123 *	1-12
A	YE, S. ET AL.: "MATRIX METALLOPROTEINASES: ..." CLINICAL SCIENCE, vol. 94, February 1998 (1998-02), pages 103-110, XP000857426 * whole disclosure *	1-12
	---	
	-/--	

☒ Further documents are listed in the continuation of box C.

☐ Patent family members are listed in annex.

### \* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

27 January 2000

Date of mailing of the international search report

03.05.00

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax (+31-70) 340-3016

Authorized officer

HERMANN R.

# INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 99/23315

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>DATABASE EMBL/GENBANK [Online]  ID HSZZ74011; AC AA368885,  18 April 1997 (1997-04-18)  ADAMS, M.D. ET AL.: "EST80295 Placenta I  homo sapiens cDNA"  XP002128855  abstract</p> <p>---</p>	2-8
X	<p>DATABASE EMBL/GENBANK [Online]  ID HSW4574; AC W92457,  18 July 1996 (1996-07-18)  HILLIER, L. ET AL.: "EST; Soares fetal  heart NbHH19W homo sapiens cDNA clone"  XP002128856  abstract</p> <p>-----</p>	2-8

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 99/ 23315

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.: 10-12  
because they relate to subject matter not required to be searched by this Authority, namely:  
Although claims 10-12 are directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.
2. ☒ Claims Nos.: 2,7  
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:  
See FURTHER INFORMATION sheet PCT/ISA/210
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

See additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-3,5-7,9,10,12 (all partially)

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Claims Nos.: 2,7

Claims 2 and 7 are -inter alia- drafted to compounds comprising as few as 18 nucleotides / 6 amino acids, which are neither defined by their exact structure, not by their exact location in the parent molecule.

Moreover, claim 2 covers

- polynucleotides of completely undefined length that hybridize to said short sequence (or the parent), and
- sequences which are complementary to the almost undefined sequences mentioned above.

Said vague structural definitions are not sufficient for a reasonable search.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

1. Claims: 1-3,5-7,9,10,12 (all partially)

A polynucleotide having the SEQ ID NO 1; subject-matter of claims 2,3,5,6,9,12 relating to said polynucleotide

2. Claims: 1-12, all partially

Polypeptide with the SEQ ID NO. 21, encoding polynucleotides, and related subject-matter

3. Claims: 1-12, all partially

Polypeptide with the SEQ ID NO. 22, encoding polynucleotides, and related subject-matter

4. Claims: 1-12, all partially

Polypeptide with the SEQ ID NO. 23, encoding polynucleotides, and related subject-matter

5. Claims: 1-3,5-7,9,10,12 (all partially)

INVENTIONS 5-20:  
Polynucleotides having the SEQ ID NOs. 3-5,7-10,12-20; subject-matter of claims 2,3,5,6,9,12 relating to said polynucleotide.

6. Claims: 1,3 (partially)

INVENTION 21:  
Further polynucleotides and encoded polypeptides of claim 1 (PLEASE NOTE: This subject has been added for completeness only. It is considered as being at least partially anticipated (infra); Further non-unity objections may arise upon payment for subject 21)

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>7</sup> :</b> <b>C07K 14/00</b>	<b>A2</b>	<b>(11) International Publication Number:</b> <b>WO 00/21986</b> <b>(43) International Publication Date:</b> 20 April 2000 (20.04.00)
<b>(21) International Application Number:</b> PCT/US99/23315 <b>(22) International Filing Date:</b> 6 October 1999 (06.10.99) <b>(30) Priority Data:</b> 09/169,289 9 October 1998 (09.10.98) US <b>(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application</b> US 09/169,289 (CIP) Filed on 9 October 1998 (09.10.98) <b>(71) Applicant (for all designated States except US):</b> INCYTE PHARMACEUTICALS, INC. [US/US]; 3174 Porter Drive, Palo Alto, CA 94304 (US). <b>(72) Inventors; and</b> <b>(75) Inventors/Applicants (for US only):</b> WALKER, Michael, G. [CA/US]; Unit 80, 1050 Borregas Avenue, Sunnyvale, CA 94089 (US). VOLKMUTH, Wayne [US/US]; 783 Roble Avenue, #1, Menlo Park, CA 94025 (US). KLINGLER, Tod, M. [US/US]; 28 Dover Court, San Carlos, CA 94070 (US). <b>(74) Agents:</b> BILLINGS, Lucy, J. et al.; Incyte Pharmaceuticals, Inc., 3174 Porter Drive, Palo Alto, CA 94304 (US).		<b>(81) Designated States:</b> AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).  <b>Published</b> <i>Without international search report and to be republished upon receipt of that report.</i>
<b>(54) Title:</b> MATRIX-REMODELING GENES  <b>(57) Abstract</b>  The invention provides novel matrix-remodeling genes and polypeptides encoded by those genes. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating or preventing diseases associated with matrix remodeling.		



**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

## MATRIX-REMODELING GENES

### TECHNICAL FIELD

5           The invention relates to novel matrix-remodeling genes identified by their coexpression with known matrix-remodeling genes. The invention also relates to the use of these biomolecules in diagnosis, prognosis, prevention, treatment, and evaluation of therapies for diseases, particularly diseases associated with matrix-remodeling such as cancer, cardiomyopathy, arthritis, angiogenesis, diabetic necrosis, atherosclerosis, fibrosis, and  
10   ulceration.

### BACKGROUND OF THE INVENTION

Matrix remodeling is associated with the construction, destruction, and reorganization of extracellular matrix components and is essential in normal cellular functions and also in many disease processes. These disease processes include metastatic cancer, cardiomyopathy, arthritis, angiogenesis, diabetic necrosis, atherosclerosis, fibrosis, and ulceration (Alexander  
15   and Werb (1991) In: Cell Biology of Extracellular Matrix, Plenum Press, New York NY, pp. 255-302; Schuppan et al. (1993) In: Extracellular Matrix, Marcel Dekker, New York NY, pp. 201-254; Zvibel and Kraft (1993) In: Extracellular Matrix, Marcel Dekker, New York NY, pp. 559-580; Shanahan et al. (1994) J Clin Invest 93:2393-402; Kielty and Shuttleworth (1995) Int  
20   J Biochem Cell Biol 27:747-60; Bitar and Labbad (1996) J Surg Res 61:113-9; Dourado et al. (1996) Osteoarthritis Cartilage 4:187-96; Grant et al. (1996) Regul. Pept. 67:137-44; Gunja-Smith et al. (1996) Am J Pathol 148:1639-48; Alcolado et al. (1997) Clin. Sci 92:103-12; Cs-Szabo et al. (1997) Arthritis Rheum 40:1037-45; Hayward and Brock (1997) Hum Mutat 10:415-23; Ledda et al. (1997) J Invest Dermatol 108:210-4; Hayashido et al. (1998) Int J  
25   Cancer 75:654-8; Ito et al. (1998) Kidney Int 53:853-61; Nelson et al. (1998) Cancer Res 58:232-6).

Many genes that participate in and regulate matrix remodeling are known, but many remain to be identified. Identification of currently unknown genes will provide new diagnostic and therapeutic targets. In addition, these genes will provide new opportunities for  
30   therapeutic tissue engineering--the use of drugs or biologicals to direct the creation of new tissues such as skin, pancreas, or liver that can replace tissues lost to disease or trauma.

The present invention provides new compositions that are useful for diagnosis, prognosis, treatment, prevention, and evaluation of therapies for diseases associated with matrix remodeling. We have implemented a method for analyzing gene expression patterns and have identified 20 novel matrix-remodeling genes by their coexpression with known  
5 matrix-remodeling genes.

### SUMMARY OF THE INVENTION

In one aspect, the invention provides for a substantially purified polynucleotide comprising a gene that is coexpressed with one or more known matrix-remodeling genes in a plurality of biological samples. Preferably, each known matrix-remodeling gene is selected  
10 from the group consisting of osteonectin (BM-40), chondroitin/dermatan sulfate proteoglycans (C/DSPG), collagen I, II, III, and IV, connective tissue growth factor (CTGF), fibrillin, fibronectins, fibronectin receptor (fibr-r), fibulin 1, heparan sulfate proteoglycans (HSPG), extracellular matrix protein (hevin), insulin-like growth factor 1 (IGF 1), insulin-like growth factor binding protein (IGFBP), laminin, lumican, matrix Gla protein (MGP), matrix  
15 metalloproteases (MMP), and tissue inhibitors of matrix metalloproteinase 1, 2, and 3 (TIMP 1, 2, and 3). Preferred embodiments are (a) a polynucleotide sequence selected from the group consisting of SEQ ID NOs:1-20; (b) a polynucleotide sequence which encodes a polypeptide sequence of SEQ ID NOs:21, 22, and 23; (c) a polynucleotide sequence having at least 70% identity to the polynucleotide sequence of (a) or (b); (d) a polynucleotide  
20 sequence comprising at least 18 sequential nucleotides of the polynucleotide sequence of (a), (b), or (c); (e) a polynucleotide which hybridizes under stringent conditions to the polynucleotide of (a), (b), (c), or (d); or (f) a polynucleotide sequence which is complementary to the polynucleotide sequence of (a), (b), (c), (d) or (e). Furthermore, the invention provides an expression vector comprising any of the above described  
25 polynucleotides and host cells comprising the expression vector. Still further, the invention provides a method for treating or preventing a disease or condition associated with the altered expression of a gene that is coexpressed with one or more known matrix-remodeling genes in a sample comprising administering to a subject in need the above-described polynucleotides in an amount effective for treating or preventing said disease.

30 In a second aspect, the invention provides a substantially purified polypeptide comprising the gene product of a gene that is coexpressed with one or more known matrix-

remodeling genes in a plurality of biological samples. The known matrix-remodeling gene may be selected from the group consisting of osteonectin (BM-40), chondroitin/dermatan sulfate proteoglycans (C/DSPG), collagen I, II, III, and IV, connective tissue growth factor (CTGF), fibrillin, fibronectins, fibronectin receptors (fibr-r), fibulin 1, heparan sulfate

5 proteoglycans (HSPG), extracellular matrix protein (hevin), insulin-like growth factor 1 (IGF 1), insulin-like growth factor binding protein (IGFBP), laminin, lumican, matrix Gla protein (MGP), matrix metalloproteases (MMP), and tissue inhibitors of matrix metalloproteinase 1, 2, and 3 (TIMP 1, 2, and 3). Preferred embodiments are polypeptides comprising (a) the polypeptide sequence of SEQ ID NO:21, 22, or 23; (b) a polypeptide sequence having at least

10 85% identity to the polypeptide sequence of (a); and (c) a polypeptide sequence comprising at least 6 sequential amino acids of the polypeptide sequence of (a) or (b). Additionally, the invention provides antibodies that bind specifically to any of the above described polypeptides and a method for treating or preventing a disease or condition associated with the altered expression of a gene that is coexpressed with one or more known matrix-remodeling genes in

15 a sample comprising administering to a subject in need such an antibody in an amount effective for treating or preventing said disease.

In another aspect, the invention provides a pharmaceutical composition comprising the polynucleotide of claim 2 or the polypeptide of claim 3 in conjunction with a suitable pharmaceutical carrier or a method for treating or preventing a disease or condition associated

20 with the altered expression of a gene that is coexpressed with one or more known matrix-remodeling genes in a sample comprising administering to a subject in need such composition in an amount effective for treating or preventing said disease.

In yet a further aspect, the invention provides a method for diagnosing a disease or condition associated with the altered expression of a gene that is coexpressed with one or more

25 known matrix-remodeling genes in a sample, wherein each known matrix-remodeling gene is selected from the group consisting of osteonectin (BM-40), chondroitin/dermatan sulfate proteoglycans (C/DSPG), collagen I, II, III, and IV, connective tissue growth factor (CTGF), fibrillin, fibronectins, fibronectin receptor (fibr-r), fibulin 1, heparan sulfate proteoglycans (HSPG), extracellular matrix protein (hevin), insulin-like growth factor 1 (IGF 1), insulin-like

30 growth factor binding protein (IGFBP), laminin, lumican, matrix Gla protein (MGP), matrix metalloproteases (MMP), and tissue inhibitors of matrix metalloproteinase 1, 2, and 3 (TIMP

1, 2, and 3). The method comprises the steps of (a) providing the sample comprising one of more of said coexpressed genes; (b) hybridizing the polynucleotide of the coexpressed genes under conditions effective to form one or more hybridization complexes; and (c) detecting the hybridization complexes, wherein the altered level of one or more of the hybridization  
5 complexes in a diseased sample compared with the level of hybridization complexes in a non-diseased sample correlates with the presence of the disease or condition in the sample.

### BRIEF DESCRIPTION OF THE SEQUENCE LISTING

The Sequence Listing provides exemplary matrix-remodeling-associated sequences including polynucleotide sequences, SEQ ID NOs:1-20, and polypeptide sequences, SEQ ID  
10 NOs:21-23. Each sequence is identified by a sequence identification number (SEQ ID NO) and by the Incyte Clone number from which the sequence was first identified.

### DESCRIPTION OF THE INVENTION

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include the plural reference unless the context clearly dictates otherwise.  
15 Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

### DEFINITIONS

"NSEQ" refers generally to a polynucleotide sequence of the present invention,  
20 including SEQ ID NOs:1-20. "PSEQ" refers generally to a polypeptide sequence of the present invention, including SEQ ID NOs:21-23.

A "variant" refers to either a polynucleotide or a polypeptide whose sequence diverges from SEQ ID NOs:1-20 or SEQ ID NOs:21-23, respectively. Polynucleotide sequence divergence may result from mutational changes such as deletions, additions, and substitutions  
25 of one or more nucleotides; it may also occur because of differences in codon usage. Each of these types of changes may occur alone, or in combination, one or more times in a given sequence. Polypeptide variants include sequences that possess at least one structural or functional characteristic of SEQ ID NOs:21-23.

A "fragment" can refer to a nucleic acid sequence that is preferably at least 20 nucleic  
30 acids in length, more preferably 40 nucleic acids, and most preferably 60 nucleic acids in length, and encompasses, for example, fragments consisting of nucleic acids 1-50 or 200-500

of SEQ ID NOs:1-20. A "fragment" can also refer to polypeptide sequences which are preferably at least 5 to about 15 amino acids in length, most preferably at least 10 amino acids long, and which retain some biological activity or immunological activity of, for example, a sequence selected from SEQ ID NOs:21-23.

- 5 "Gene" or "gene sequence" refers to the partial or complete coding sequence of a transcript. The term also refers to sequences corresponding to 5' or 3' untranslated regions or 5' or 3' untranslated regions including partial or complete coding sequences of a gene. Typically, the novel gene sequences may or may not be homologous to annotated sequences found in public or private databases. The gene may be in a sense or antisense  
10 (complementary) orientation.

- "Known matrix-remodeling gene" refers to a gene sequence which has been previously identified as useful in the diagnosis, treatment, prognosis, or prevention of diseases associated with matrix remodeling. Typically, this means that the known matrix-remodeling gene is expressed at higher levels in tissue abundant in known matrix-remodeling transcripts when  
15 compared with other tissue.

- "Matrix-remodeling gene" refers to a gene sequence whose expression pattern is similar to that of the known matrix-remodeling genes and which are useful in the diagnosis, treatment, prognosis, or prevention of diseases associated with matrix remodeling. The gene sequences can also be used in the evaluation of therapies for cancer.

- 20 "Substantially purified" refers to a nucleic acid or an amino acid sequence that is removed from its natural environment and is isolated or separated, and is at least about 60% free, preferably about 75% free, and most preferably about 90% free from other components with which it is naturally present.

## THE INVENTION

- 25 The present invention encompasses a method for identifying biomolecules that are associated with a specific disease, regulatory pathway, subcellular compartment, cell type, tissue type, or species. In particular, the method identifies gene sequences useful in diagnosis, prognosis, treatment, prevention, and evaluation of therapies for diseases associated with matrix-remodeling, particularly, cancer, cardiomyopathy, arthritis, angiogenesis, diabetic  
30 necrosis, atherosclerosis, fibrosis, and ulceration.

The method provides first identifying polynucleotides that are expressed in a plurality

of cDNA libraries. The identified polynucleotides include genes of known function, genes known to be specifically expressed in a specific disease process, subcellular compartment, cell type, tissue type, or species. Additionally, the polynucleotides include genes of unknown function. The expression patterns of the known genes are then compared with those of the  
5 genes of unknown function to determine whether a specified coexpression probability threshold is met. Through this comparison, a subset of the polynucleotides for unknown function genes having a high coexpression probability with the known genes can be identified. The high coexpression probability correlates with a particular coexpression probability threshold which is less than 0.001, and more preferably less than 0.00001.

10 The polynucleotides originate from cDNA libraries derived from a variety of sources including, but not limited to, eukaryotes such as human, mouse, rat, dog, monkey, plant, and yeast and prokaryotes such as bacteria and viruses. These polynucleotides can also be selected from a variety of sequence types including, but not limited to, expressed sequence tags (ESTs), assembled polynucleotide sequences, full length gene coding regions, introns,  
15 regulatory sequences, 5' untranslated regions, and 3' untranslated regions. To have statistically significant analytical results, the polynucleotides need to be expressed in at least three cDNA libraries.

The cDNA libraries used in the coexpression analysis of the present invention can be obtained from blood vessels, heart, blood cells, cultured cells, connective tissue, epithelium,  
20 islets of Langerhans, neurons, phagocytes, biliary tract, esophagus, gastrointestinal system, liver, pancreas, fetus, placenta, chromaffin system, endocrine glands, ovary, uterus, penis, prostate, seminal vesicles, testis, bone marrow, immune system, cartilage, muscles, skeleton, central nervous system, ganglia, neuroglia, neurosecretory system, peripheral nervous system, bronchus, larynx, lung, nose, pleurus, ear, eye, mouth, pharynx, exocrine glands, bladder,  
25 kidney, ureter, and the like. The number of cDNA libraries selected can range from as few as 20 to greater than 10,000. Preferably, the number of the cDNA libraries is greater than 500.

In a preferred embodiment, gene sequences are assembled to reflect related sequences, such as assembled sequence fragments derived from a single transcript. Assembly of the polynucleotide sequences can be performed using sequences of various types including, but  
30 not limited to, ESTs, extensions, or shotgun sequences. In a most preferred embodiment, the polynucleotide sequences are derived from human sequences that have been assembled using

the algorithm disclosed in "Database and System for Storing, Comparing and Displaying Related Biomolecular Sequence Information", Lincoln et al., Serial No:60/079,469, filed March 26, 1998, incorporated herein by reference.

Experimentally, differential expression of the polynucleotides can be evaluated by methods including, but not limited to, differential display by spatial immobilization or by gel electrophoresis, genome mismatch scanning, representational difference analysis, and transcript imaging. Additionally, differential expression can be assessed by microarray technology. These methods may be used alone or in combination.

Known matrix-remodeling genes can be selected based on the use of the genes as diagnostic or prognostic markers or as therapeutic targets for diseases associated with matrix remodeling, such as cancer, cardiomyopathy, arthritis, angiogenesis, diabetic necrosis, atherosclerosis, fibrosis, and ulceration. Preferably, the known matrix-remodeling genes include osteonectin (BM-40), chondroitin/dermatan sulfate proteoglycans (C/DSPG), collagen I, II, III, and IV, connective tissue growth factor (CTGF), fibrillin, fibronectins, fibronectin receptor (fibr-r), fibulin 1, heparan sulfate proteoglycans (HSPG), extracellular matrix protein (hevin), insulin-like growth factor 1 (IGF 1), insulin-like growth factor binding protein (IGFBP), laminin, lumican, matrix Gla protein (MGP), matrix metalloproteinases (MMP), tissue inhibitors of matrix metalloproteinase 1, 2, and 3 (TIMP 1, 2, and 3), and the like.

The procedure for identifying novel genes that exhibit a statistically significant coexpression pattern with known matrix-remodeling genes is as follows. First, the presence or absence of a gene sequence in a cDNA library is defined: a gene is present in a cDNA library when at least one cDNA fragment corresponding to that gene is detected in a cDNA sample taken from the library, and a gene is absent from a library when no corresponding cDNA fragment is detected in the sample.

Second, the significance of gene coexpression is evaluated using a probability method to measure a due-to-chance probability of the coexpression. The probability method can be the Fisher exact test, the chi-squared test, or the kappa test. These tests and examples of their applications are well known in the art and can be found in standard statistics texts (Agresti, A (1990) Categorical Data Analysis, John Wiley & Sons, New York NY; Rice, JA (1988) Mathematical Statistics and Data Analysis, Duxbury Press, Pacific Grove CA). A Bonferroni



correction (Rice, *supra*, page 384) can also be applied in combination with one of the probability methods for correcting statistical results of one gene versus multiple other genes. In a preferred embodiment, the due-to-chance probability is measured by a Fisher exact test, and the threshold of the due-to-chance probability is set to less than 0.001, more preferably less than 0.00001.

To determine whether two genes, A and B, have similar coexpression patterns, occurrence data vectors can be generated as illustrated in Table 1, wherein a gene's presence is indicated by a one and its absence by a zero. A zero indicates that the gene did not occur in the library, and a one indicates that it occurred at least once.

Table 1. Occurrence data for genes A and B

	Library 1	Library 2	Library 3	...	Library N
gene A	1	1	0	...	0
gene B	1	0	1	...	0

For a given pair of genes, the occurrence data in Table 1 can be summarized in a 2 x 2 contingency table.

Table 2. Contingency table for co-occurrences of genes A and B

	Gene A present	Gene A absent	Total
Gene B present	8	2	10
Gene B absent	2	18	20
Total	10	20	30

Table 2 presents co-occurrence data for gene A and gene B in a total of 30 libraries. Both gene A and gene B occur 10 times in the libraries. Table 2 summarizes and presents 1) the number of times gene A and B are both present in a library, 2) the number of times gene A and B are both absent in a library, 3) the number of times gene A is present while gene B is absent, and 4) the number of times gene B is present while gene A is absent. The upper left entry is the number of times the two genes co-occur in a library, and the middle right entry is the number of times neither gene occurs in a library. The off diagonal entries are the number of times one gene occurs while the other does not. Both A and B are present eight times and absent 18 times, gene A is present while gene B is absent two times, and gene B is present while gene A

is absent two times. The probability ("p-value") that the above association occurs due to chance as calculated using a Fisher exact test is 0.0003. Associations are generally considered significant if a p-value is less than 0.01 (Agresti, *supra*; Rice, *supra*).

This method of estimating the probability for coexpression of two genes makes several  
5 assumptions. The method assumes that the libraries are independent and are identically sampled. However, in practical situations, the selected cDNA libraries are not entirely independent because more than one library may be obtained from a single patient or tissue, and they are not entirely identically sampled because different numbers of cDNAs may be sequenced from each library (typically ranging from 5,000 to 10,000 cDNAs per library). In  
10 addition, because a Fisher exact coexpression probability is calculated for each gene versus 41,419 other genes, a Bonferroni correction for multiple statistical tests is necessary.

Using the method of the present invention, we have identified 20 novel genes that exhibit strong association, or coexpression, with known genes that are matrix-remodeling-specific. These known matrix-remodeling genes include osteonectin (BM-40),  
15 chondroitin/dermatan sulfate proteoglycans (C/DSPG), collagen I, II, III, and IV, connective tissue growth factor (CTGF), fibrillin, fibronectins, fibronectin receptor (fibr-r), fibulin 1, heparan sulfate proteoglycans (HSPG), extracellular matrix protein (hevin), insulin-like growth factor 1 (IGF 1), insulin-like growth factor binding protein (IGFBP), laminin, lumican, matrix Gla protein (MGP), matrix metalloproteases (MMP), and tissue inhibitors of matrix  
20 metalloproteinase 1, 2, and 3 (TIMP 1, 2, and 3). The results presented in Tables 5 and 6 show that the expression of the 20 novel genes have direct or indirect association with the expression of known matrix-remodeling genes. Therefore, the novel genes can potentially be used in diagnosis, treatment, prognosis, or prevention of diseases associated with matrix remodeling, or in the evaluation of therapies for diseases associated with matrix remodeling.  
25 Further, the gene products of the 20 novel genes are potential therapeutic proteins and targets of therapeutics against diseases associated with matrix remodeling.

Therefore, in one embodiment, the present invention encompasses a polynucleotide sequence comprising the sequence of SEQ ID NOs:1-20. These 20 polynucleotides are shown  
by the method of the present invention to have strong coexpression association with known  
30 matrix-remodeling genes and with each other. The invention also encompasses a variant of the polynucleotide sequence, its complement, or 18 consecutive nucleotides of a sequence

provided in the above described sequences. Variant polynucleotide sequences typically have at least about 70%, more preferably at least about 85%, and most preferably at least about 95% polynucleotide sequence identity to NSEQ.

One preferred method for identifying variants entails using NSEQ and/or PSEQ sequences to search against the GenBank primate (pri), rodent (rod), and mammalian (mam), vertebrate (vrtp), and eukaryote (eukp) databases, SwissProt, BLOCKS (Bairoch et al. (1997) Nucleic Acids Res 25:217-221), PFAM, and other databases that contain previously identified and annotated motifs, sequences, and gene functions. Methods that search for primary sequence patterns with secondary structure gap penalties (Smith et al. (1992) Protein Engineering 5:35-51) as well as algorithms such as BLAST (Basic Local Alignment Search Tool; Altschul (1993) J Mol Evol 36:290-300; and Altschul et al. (1990) J Mol Biol 215:403-410), BLOCKS (Henikoff and Henikoff (1991) Nucleic Acids Res 19:6565-6572), Hidden Markov Models (HMM; Eddy (1996) Cur Opin Str Biol 6:361-365; Sonnhammer et al. (1997) Proteins 28:405-420), and the like, can be used to manipulate and analyze nucleotide and amino acid sequences. These databases, algorithms and other methods are well known in the art and are described in Ausubel et al. (1997; Short Protocols in Molecular Biology, John Wiley & Sons, New York NY) and in Meyers (1995; Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853).

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to SEQ ID NOS:1-20, and fragments thereof under stringent conditions. Stringent conditions can be defined by salt concentration, temperature, and other chemicals and conditions well known in the art. In particular, stringency can be increased by reducing the concentration of salt, or raising the hybridization temperature. Varying additional parameters, such as hybridization time, the concentration of detergent or solvent, and the inclusion or exclusion of carrier DNA, are well known to those skilled in the art. Additional variations on these conditions will be readily apparent to those skilled in the art (Wahl and Berger (1987) Methods Enzymol 152:399-407; Kimmel (1987) Methods Enzymol 152:507-511; Ausubel supra; and Sambrook et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY).

NSEQ or the polynucleotide sequences encoding PSEQ can be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to

detect upstream sequences, such as promoters and regulatory elements. (See, e.g., Dieffenbach and Dveksler (1995) PCR Primer, a Laboratory Manual, Cold Spring Harbor Press, Plainview NY; Sarkar (1993) PCR Methods Applic 2:318-322; Triglia et al. (1988) Nucleic Acids Res 16:8186; Lagerstrom et al. (1991) PCR Methods Applic 1:111-119; and  
5 Parker et al. (1991) Nucleic Acids Res 19:3055-306). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto, CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 Primer Analysis software (National Biosciences,  
10 Plymouth MN) or another appropriate program, to be about 18 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

In another aspect of the invention, NSEQ or the polynucleotide sequences encoding PSEQ can be cloned in recombinant DNA molecules that direct expression of PSEQ or the  
15 polypeptides encoded by NSEQ, or structural or functional fragments thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express the polypeptides of PSEQ or the polypeptides encoded by NSEQ. The nucleotide sequences of the present invention can be engineered using methods  
20 generally known in the art in order to alter the nucleotide sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce  
25 mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

In order to express a biologically active polypeptide encoded by NSEQ, NSEQ or the polynucleotide sequences encoding PSEQ, or derivatives thereof, may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for  
30 transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible

promoters, and 5' and 3' untranslated regions in the vector and in NSEQ or polynucleotide sequences encoding PSEQ. Methods which are well known to those skilled in the art may be used to construct expression vectors containing NSEQ or polynucleotide sequences encoding PSEQ and appropriate transcriptional and translational control elements. These methods  
5 include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook (supra) and Ausubel (supra)).

A variety of expression vector/host cell systems may be utilized to contain and express NSEQ or polynucleotide sequences encoding PSEQ. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or  
10 cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (baculovirus); plant cell systems transformed with viral expression vectors, cauliflower mosaic virus (CaMV) or tobacco mosaic virus (TMV), or with bacterial expression vectors (Ti or pBR322 plasmids); or animal cell systems. The invention is not limited by the host cell employed. For long term production of  
15 recombinant proteins in mammalian systems, stable expression of a polypeptide encoded by NSEQ in cell lines is preferred. For example, NSEQ or sequences encoding PSEQ can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector.

20 In general, host cells that contain NSEQ and that express PSEQ may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences. Immunological  
25 methods for detecting and measuring the expression of PSEQ using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS).

Host cells transformed with NSEQ or polynucleotide sequences encoding PSEQ may  
30 be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly

depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides of NSEQ or polynucleotides encoding PSEQ may be designed to contain signal sequences which direct secretion of PSEQ or polypeptides encoded by NSEQ through a prokaryotic or eukaryotic cell membrane.

5 In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" form of the protein may also be used to specify protein targeting, folding,  
10 and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38), are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant NSEQ or  
15 nucleic acid sequences encoding PSEQ are ligated to a heterologous sequence resulting in translation of a fusion protein containing heterologous protein moieties in any of the aforementioned host systems. Such heterologous protein moieties facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin  
20 (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, hemagglutinin (HA) and monoclonal antibody epitopes..

In another embodiment, NSEQ or sequences encoding PSEQ are synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers et al. (1980) Nucleic Acids Symp Ser (7) 215-223; Horn et al. (1980) Nucleic Acids Symp Ser (7) 225-232;  
25 and Ausubel, supra). Alternatively, PSEQ or a polypeptide sequence encoded by NSEQ itself, or a fragment thereof, may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solid-phase techniques (Roberge et al. (1995) Science 269:202-204). Automated synthesis may be achieved using the ABI 431A Peptide synthesizer (PE Biosystems, Foster City CA). Additionally, PSEQ or the amino acid  
30 sequence encoded by NSEQ, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a polypeptide

variant.

In another embodiment, the invention provides a substantially purified polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23 or fragments thereof.

## 5 **DIAGNOSTICS and THERAPEUTICS**

The sequences of these genes can be used in diagnosis, prognosis, treatment, prevention, and evaluation of therapies for diseases associated with matrix-remodeling, particularly cancer, cardiomyopathy, arthritis, angiogenesis, diabetic necrosis, atherosclerosis, fibrosis, and ulceration. Further, the amino acid sequences encoded by the novel genes are  
10 potential therapeutic proteins and targets of anti-cancer therapeutics or for the treatment of other diseases associated with matrix remodeling.

In one preferred embodiment, the polynucleotide sequences of NSEQ or the polynucleotides encoding PSEQ are used for diagnostic purposes to investigate the altered expression of PSEQ, and to monitor regulation of the levels of mRNA or the polypeptides  
15 encoded by NSEQ during therapeutic intervention. The polynucleotides may be at least 18 nucleotides long, and may be complementary RNA or DNA molecules, branched nucleic acids, or peptide nucleic acids (PNAs). Alternatively, the polynucleotides are used to detect and quantitate gene expression in samples in which expression of PSEQ or the polypeptides encoded by NSEQ are correlated with disease. Additionally, NSEQ or the polynucleotides  
20 encoding PSEQ can be used to detect genetic polymorphisms associated with a disease. These polymorphisms may be detected at the transcript cDNA or genomic level.

The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification (maximal, high, intermediate, or low), will determine  
25 whether the probe identifies only naturally occurring sequences encoding PSEQ, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and should preferably have at least 70% sequence identity to any of the NSEQ or PSEQ-encoding sequences.

Means for producing specific hybridization probes for DNAs encoding PSEQ include  
30 the cloning of NSEQ or polynucleotide sequences encoding PSEQ into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available,

and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as  $^{32}\text{P}$  or  $^{35}\text{S}$ , or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin  
5 coupling systems, by fluorescent labels and the like. The polynucleotide sequences encoding PSEQ may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; and in microarrays utilizing fluids or tissues from patients to detect altered NSEQ expression. Such qualitative or quantitative methods are well known in the art.

10 NSEQ or the nucleotide sequences encoding PSEQ can be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantitated and compared with a standard value, typically, derived from a non-diseased sample. If the amount of signal in the patient sample is significantly  
15 altered in comparison to the standard value then the presence of altered levels of nucleotide sequences of NSEQ and those encoding PSEQ in the sample indicates the presence of the associated disease. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

20 Once the presence of a disease is established and a treatment protocol is initiated, hybridization or amplification assays can be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in a healthy subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

25 The polynucleotides may be used for the diagnosis of a variety of diseases associated with matrix-remodeling including cancer such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary  
30 glands, skin, spleen, testis, thymus, thyroid, and uterus, cardiomyopathy, arthritis, angiogenesis, diabetic necrosis, atherosclerosis, fibrosis, and ulceration.



Alternatively, the polynucleotides may be used as targets in a microarray. The microarray can be used to monitor the expression level of large numbers of genes simultaneously and to identify splice variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disease, to diagnose a disease, and to develop and monitor the activities of therapeutic agents.

In yet another alternative, polynucleotides may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Fluorescent in situ hybridization (FISH) may be correlated with other physical chromosome mapping techniques and genetic map data. (See, e.g., Heinz-Ulrich et al. (1995) in Meyers, supra, pp. 965-968.)

10 In another embodiment, antibodies which specifically bind PSEQ may be used for the diagnosis of diseases characterized by the over-or-underexpression of PSEQ or polypeptides encoded by NSEQ. A variety of protocols for measuring PSEQ or the polypeptides encoded by NSEQ, including ELISAs, RIAs, and FACS, are well known in the art and provide a basis for diagnosing altered or abnormal levels of the expression of PSEQ or the polypeptides  
15 encoded by NSEQ. Standard values for PSEQ expression are established by combining body fluids or cell extracts taken from healthy subjects, preferably human, with antibody to PSEQ or a polypeptide encoded by NSEQ under conditions suitable for complex formation. The amount of complex formation may be quantitated by various methods, preferably by photometric means. Quantities of PSEQ or the polypeptides encoded by NSEQ expressed in  
20 disease samples from, for example, biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing or monitoring disease. Alternatively, one may use competitive drug screening assays in which neutralizing antibodies capable of binding PSEQ or the polypeptides encoded by NSEQ specifically compete with a test compound for binding the polypeptides. Antibodies can be  
25 used to detect the presence of any peptide which shares one or more antigenic determinants with PSEQ or the polypeptides encoded by NSEQ.

In another aspect, the polynucleotides and polypeptides of the present invention can be employed for treatment or the monitoring of therapeutic treatments for cancer. The polynucleotides of NSEQ or those encoding PSEQ, or any fragment or complement thereof,  
30 may be used for therapeutic purposes. In one aspect, the complement of the polynucleotides of NSEQ or those encoding PSEQ may be used in situations in which it would be desirable to

block the transcription or translation of the mRNA.

Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. Methods which are well known to those skilled in the art can be used to construct vectors to express nucleic acid sequences complementary to the polynucleotides encoding PSEQ. (See, e.g., Sambrook, supra; and Ausubel, supra.)

Genes having polynucleotide sequences of NSEQ or those encoding PSEQ can be turned off by transforming a cell or tissue with expression vectors which express high levels of a polynucleotide, or fragment thereof, encoding PSEQ. Such constructs may be used to introduce untranslatable sense or antisense sequences into a cell. Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, are preferred. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee et al. (1994) In: Huber and Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. Alternatively, nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases may be included.

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g.,

Goldman et al. (1997) Nature Biotechnology 15:462-466.)

Further, an antagonist or antibody of a polypeptide of PSEQ or encoded by NSEQ may be administered to a subject to treat or prevent a cancer associated with increased expression or activity of PSEQ. An antibody which specifically binds the polypeptide may be used  
5 directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express the the polypeptide.

Antibodies to PSEQ or polypeptides encoded by NSEQ may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments  
10 produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are especially preferred for therapeutic use. Monoclonal antibodies to PSEQ may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. In  
15 addition, techniques developed for the production of chimeric antibodies can be used. (See, for example, Meyers, supra.) Alternatively, techniques described for the production of single chain antibodies may be employed. Antibody fragments which contain specific binding sites for PSEQ or the polypeptide sequences encoded by NSEQ may also be generated.

Various immunoassays may be used for screening to identify antibodies having the  
20 desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art.

Yet further, an agonist of a polypeptide of PSEQ or that encoded by NSEQ may be administered to a subject to treat or prevent a cancer associated with decreased expression or  
25 activity of the polypeptide.

An additional aspect of the invention relates to the administration of a pharmaceutical or sterile composition, in conjunction with a pharmaceutically acceptable carrier, for any of the therapeutic effects discussed above. Such pharmaceutical compositions may consist of polypeptides of PSEQ or those encoded by NSEQ, antibodies to the polypeptides, and  
30 mimetics, agonists, antagonists, or inhibitors of the polypeptides. The compositions may be administered alone or in combination with at least one other agent, such as a stabilizing

compound, which may be administered in any sterile, biocompatible pharmaceutical carrier including, but not limited to, saline, buffered saline, dextrose, and water. The compositions may be administered to a patient alone, or in combination with other agents, drugs, or hormones.

5       The pharmaceutical compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

10       In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically-acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA).

15       For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells or in animal models such as mice, rats, rabbits, dogs, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

20       A therapeutically effective dose refers to that amount of active ingredient, for example, polypeptides of PSEQ or those encoded by NSEQ, or fragments thereof, antibodies of the polypeptides, and agonists, antagonists or inhibitors of the polypeptides, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by  
25       calculating the  $ED_{50}$  (the dose therapeutically effective in 50% of the population) or  $LD_{50}$  (the dose lethal to 50% of the population) statistics.

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as dogs, cats, cows, horses, rabbits, monkeys, and most preferably, humans.

30

## EXAMPLES

It is understood that this invention is not limited to the particular methodology,

protocols, and reagents described, as these may vary. It is also understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims. The examples below are provide to illustrate the subject invention and are  
5 not included for the purpose of limiting the invention.

## **I cDNA Library Construction**

The cDNA library, THYMFET02, was selected to demonstrate the construction of the cDNA libraries from which novel matrix remodeling genes were derived. The THYMFET02 cDNA library was constructed from microscopically normal thymus tissue obtained from a  
10 Caucasian female fetus who died at 17 weeks gestation from anencephaly. Serology was negative; family history included tobacco abuse and gastritis.

The frozen tissue was homogenized and lysed in TRIZOL reagent (1 gm tissue/10 ml TRIZOL; Life Technologies, Rockville MD), a monoplastic solution of phenol and guanidine isothiocyanate, using a POLYTRON homogenizer (PT-3000; Brinkmann Instruments,  
15 Westbury NY). After a brief incubation on ice, chloroform was added (1:5 v/v), and the lysate was centrifuged. The upper chloroform layer was removed, and the RNA was precipitated with isopropanol, resuspended in DEPC-treated water, and treated with DNase for 25 min at 37°C. The mRNA was reextracted once with acid phenol-chloroform pH 4.7 and precipitated using 0.3 M sodium acetate and 2.5 volumes ethanol. The mRNA was isolated using the  
20 OLIGOTEX kit (Qiagen, Chatsworth CA) and used to construct the cDNA library.

The mRNA was handled according to the recommended protocols in the SUPERScript Plasmid system (Life Technologies). The cDNAs were fractionated on a SEPHAROSE CL4B column (Amersham Pharmacia Biotech, Piscataway NJ), and those cDNAs exceeding 400 bp were ligated into pINCY 1 plasmid (Incyte Pharmaceuticals, Palo  
25 Alto CA). The plasmid was subsequently transformed into DH5a competent cells (Life Technologies).

## **II Isolation and Sequencing of cDNA Clones**

Plasmid DNA was released from the cells and purified using the REAL Prep 96 Plasmid kit ( Qiagen). This kit enabled the simultaneous purification of 96 samples in a 96-  
30 well block using multi-channel reagent dispensers. The recommended protocol was employed except for the following changes: 1) the bacteria were cultured in 1 ml of sterile Terrific Broth

(Life Technologies) with carbenicillin at 25 mg/L and glycerol at 0.4%; 2) after inoculation, the cultures were incubated for 19 hours and at the end of incubation, the cells were lysed with 0.3 ml of lysis buffer; and 3) following isopropanol precipitation, the plasmid DNA pellet was resuspended in 0.1 ml of distilled water. After the last step in the protocol, samples were  
5 transferred to a 96-well block for storage at 4°C.

The cDNAs were prepared using a MICROLAB 2200 (Hamilton, Reno NV) in combination with DNA ENGINE thermal cyclers (PTC200; MJ Research, Watertown MA) and sequenced by the method of Sanger et al. (1975, J Mol Biol 94:441f) using ABI PRISM 377 DNA sequencing systems.

### 10 III Selection, Assembly, and Characterization of Sequences

The sequences used for coexpression analysis were assembled from EST sequences, 5' and 3' longread sequences, and full length coding sequences. Selected assembled sequences were expressed in at least three cDNA libraries.

The assembly process is described as follows. EST sequence chromatograms were  
15 processed and verified. Quality scores were obtained using PHRED (Ewing et al. (1998) Genome Res 8:175-185; Ewing and Green (1998) Genome Res 8:186-194). Then the edited sequences were loaded into a relational database management system (RDBMS). The EST sequences were clustered into an initial set of bins using BLAST with a product score of 50. All clusters of two or more sequences were created as bins. The overlapping sequences  
20 represented in a bin correspond to the sequence of a transcribed gene.

Assembly of the component sequences within each bin was performed using a modification of PHRAP, a publicly available program for assembling DNA fragments (Phil Green, University of Washington, Seattle WA). Bins that showed 82% identity from a local pair-wise alignment between any of the consensus sequences were merged.

25 Bins were annotated by screening the consensus sequence in each bin against public databases, such as GBpri and GenPept from NCBI. The annotation process involved a FASTn screen against the GBpri database in GenBank. Those hits with a percent identity of greater than or equal to 70% and an alignment length of greater than or equal to 100 base pairs were recorded as homolog hits. The residual unannotated sequences were screened by FASTx  
30 against GenPept. Those hits with an E value of less than or equal to  $10^{-8}$  are recorded as homolog hits.

Sequences were then reclustered using BLASTn and Cross-Match, a program for rapid protein and nucleic acid sequence comparison and database search (Green, supra), sequentially. Any BLAST alignment between a sequence and a consensus sequence with a score greater than 150 was realigned using cross-match. The sequence was added to the bin whose consensus sequence gave the highest Smith-Waterman score amongst local alignments with at least 82% identity. Non-matching sequences created new bins. The assembly and consensus generation processes were performed for the new bins.

#### IV Coexpression Analyses of Known Matrix-remodeling Genes

Twenty one known matrix-remodeling genes were selected to identify novel genes that are closely associated with matrix remodeling. The known genes were osteonectin (BM-40), chondroitin/dermatan sulfate proteoglycans (C/DSPG), collagen I, II, III, and IV (coll-I, coll-II, and coll-III), connective tissue growth factor (CTGF), fibrillin, fibronectins, fibronectin receptor (fibr-r), fibulin 1, heparan sulfate proteoglycans (HSPG), extracellular matrix protein (hevin), insulin-like growth factor 1 (IGF 1), insulin-like growth factor binding protein (IGFBP), laminin, lumican, matrix Gla protein (MGP), matrix metalloproteases (MMP), and tissue inhibitors of matrix metalloproteinase 1, 2, and 3 (TIMP 1, 2, and 3). The protein products of the known matrix-remodeling genes may be categorized as follows.

1. Extracellular matrix component protein. These proteins include collagens, proteoglycans, fibrillin, fibronectin, fibulin, and laminin that constitute the major structures of the extracellular matrix.
2. Matrix proteases and matrix protease inhibitors. These proteins include matrix metalloproteases (MMPs) such as the collagenases, and MMP inhibitors such as the tissue-inhibitors of matrix metalloproteases (TIMPs).
3. Regulatory proteins that control expression of matrix-remodeling genes. Such regulatory proteins include connective tissue growth factor, insulin-like growth factor, osteonectin (BM-40), and the receptors for and inhibitors of these proteins.

The known matrix-remodeling genes that we examined in this analysis, and brief descriptions of their functions, are listed in Table 4. Detailed descriptions of their roles in matrix remodeling may be found in the cited articles and reviews.

Table 4. Known Matrix-remodeling Genes.

Gene	Description & References
BM-40	<p>Alternate names: SPARC, osteonectin</p> <p>Regulates connective tissue remodeling, wound healing, angiogenesis</p> <p>Induces matrix metalloprotease synthesis (collagenase &amp; gelatinase)</p> <p>Regulates cell movement and proliferation</p> <p>Expression increased in neoplastic melanoma, fibrosis, angiogenesis.</p> <p>(Kamihagi et al. (1994) <i>Biochem Biophys Res Commun</i> 200:423-8; Lane et al. (1994) <i>J Cell Biol</i> 125:929-43; Inagaki et al. (1996) <i>Life Sci</i> 58:927-34; Ledda et al. (1997) <i>J Invest Dermatol</i> 108:210-4; Shankavaram et al. (1997) <i>J Cell Physiol</i> 173:327-34.)</p>
5	
C/DSPG	<p>Chondroitin/dermatan sulfate proteoglycans</p> <p>Major extracellular matrix proteoglycan</p> <p>Regulate cell proliferation, attachment and migration.</p> <p>Darnell et al. (1990) <i>Molecular Cell Biology</i>, Scientific American Press, New York NY; Toole (1991) In: <i>Cell Biology of Extracellular Matrix</i>, Plenum, New York NY, pp. 305-341; Beck et al. (1993) <i>Biochem Biophys Res Commun</i> 190:616-23)</p>
10	
Collagens	<p>Family of fibrous structural proteins (collagen I, II, III, IV, etc.)</p> <p>Most abundant structural component of the extracellular matrix</p> <p>Secreted as procollagen; converted to collagen by MMPs</p> <p>(Alexander and Werb (1991) In: <i>Cell Biology of Extracellular Matrix</i>, pp. 255-302; Adams (1993) In: <i>Extracellular Matrix</i>, Marcel Dekker, New York, NY pp. 91-119; Schuppan et al. (1993) In: <i>Extracellular Matrix</i>, pp. 201-254.)</p>
15	
CTGF	<p>Connective tissue growth factor</p> <p>Mediates induction of matrix synthesis and fibrosis</p> <p>(Grotendorst (1997) <i>Cytokine Growth Factor Rev</i> 8:171-9; Oemar and Luscher (1997) <i>Arterioscler Thromb Vasc Biol</i> 17:1483-9; Ito et al. (1998) <i>Kidney Int</i> 53:853-61.)</p>
20	
fibrillin	<p>Major component of extracellular microfibrills (matrix elastic network)</p> <p>Present in connective tissue throughout the body</p> <p>(Kielty and Shuttleworth (1995) <i>Int J Biochem Cell Biol</i> 27:747-60; Haynes et al. (1997) <i>Br J Dermatol</i> 137:17-23; Hayward and Brock (1997) <i>Hum Mutat</i> 10:415-23.)</p>
25	
fibronectins	<p>Family of extracellular matrix glycoproteins</p> <p>Anchor cells to the matrix</p> <p>Bind matrix proteins to cell surface receptors</p>
30	
fibr-r	<p>Fibronectin receptor</p> <p>Fibronectin receptors regulate cell adhesion &amp; migration</p> <p>(Darnell et al. (1990) <i>Molecular Cell Biology</i>, Scientific American Press, New York NY; Ruoslahti (1991) <i>Cell Biology of Extracellular Matrix</i>, pp. 343-363; Yamada (1991) <i>Cell Biology of Extracellular Matrix</i>, pp. 111-146.)</p>
fibulin 1	<p>Fibronectin-binding extracellular matrix protein</p>



- Mediates platelet adhesion via a bridge of fibrinogen  
 Cleaved by matrix metalloproteinases  
 Inhibits breast and ovarian cancer cell motility  
 (Argraves et al. (1990) J Cell Biol 111:3155-64; Sasaki et al. (1996) Eur J Biochem 240:427-34; Hayashido et al. (1998) Int J Cancer 75:654-8.)
- 5 HSPG Heparan sulfate proteoglycans  
 Extracellular matrix proteoglycan found on cell surface of many cell types  
 Regulate cell interactions with the extracellular matrix  
 Bind to collagens and fibronectin in the matrix  
 10 Regulate cell proliferation, attachment and migration  
 (Darnell et al. (1990) ; Toole (1991) In: Cell Biology of Extracellular Matrix, pp. 305-341; Schuppan et al. (1993) In: Extracellular Matrix, pp. 201-254.)
- hevin Extracellular matrix protein  
 Homolog to BM-40  
 15 Regulates cell adhesion and migration  
 Downregulated in metastatic prostate cancer, lung cancer  
 (Girard and Springer (1996) J Biol Chem 271:4511-7; Bendik et al. Cancer Res 58:232-6.)
- IGF 1 Insulin-like growth factor  
 Regulates matrix homeostasis and remodeling  
 20 Regulates aggregation, growth and survival of cancer cells  
 (Aston et al. (1995) Am J Respir Crit Care Med 151:1597-603; Bitar and Labbad (1996) J Surg Res 61:113-9; Guvakova and Surmacz (1997) Exp Cell Res 231:149-62; Sunic et al. (1998) Endocrinology 139:2356-62.)
- IGFBP Insulin-like growth factor binding protein  
 Regulates IGF-1 bioavailability (binds IGF-1 more strongly than the receptor)  
 25 Degraded by matrix metalloproteases  
 (Kiefer et al. (1991) Biochem Biophys Res Commun 176:219-25; Fowlkes et al. (1995) Prog Growth Factor Res 6:255-63; Parker et al. (1996) J Biol Chem 271:13523-9.)
- laminin Major protein in basal lamina, with collagen, HSPG, and entactin  
 Anchors cells to the matrix by binding collagen, HSGP and heparin  
 30 Laminins and collagens are the main targets of MMPs  
 Regulates cell attachment, migration, growth, and differentiation  
 (Yamada et al. (1993) In: Extracellular Matrix, pp. 49-66; Giannelli et al. (1997) Science 277:225-8; Quaranta and Plopper (1997) Kidney Int 51: 1441-6; Soini et al. (1997) Hum Pathol 28:220-6.)
- lumican Extracellular proteoglycan  
 35 Organizes collagen fibrils in extracellular matrix  
 (Dourado et al. (1996) Osteoarthritis Cartilage 4:187-96; Scott (1996) Biochemistry 35:8795-9; Cs-Szabo et al. (1997) Arthritis Rheum 40:1037-45.)
- MGP Matrix Gla protein  
 Regulates calcification of cartilage

Marker for osteoblast activity

(Shanahan et al. (1994) J Clin Invest 93:2393-402; Luo et al. (1997) Nature 386:78-81; Martinetti et al. (1997) Tumour Biol 18:197-205)

- 5 MMP Family of Matrix Metalloproteases (including collagenases)  
 Cleave procollagen to produce collagen  
 (Alexander and Werb (1991) In: Cell Biology of Extracellular Matrix, pp. 255-302; Adams (1993) In: Extracellular Matrix, pp. 91-119; Schuppan et al. (1993) In: Extracellular Matrix pp. 201-254.)
- 10 TIMP 1, 2, 3 Tissue inhibitors of matrix metalloproteinases  
 Bind and inactivate matrix proteases  
 (Schuppan et al. (1993) In: Extracellular Matrix, pp. 201-254; Zvibel and Kraft (1993) In: Extracellular Matrix, pp. 559-580.)

The coexpression of the 21 known genes with each other is shown in Table 5. The entries in Table 5 are the negative log of the p-value ( $-\log p$ ) for the coexpression of the two genes. As shown, the method successfully identified the strong association of the known genes among themselves, indicating that the coexpression analysis method of the present invention was effective in identifying genes that are closely associated with matrix remodeling.

Table 5. Coexpression of 21 known matrix-remodeling genes. ( $-\log p$ )

	laminin	fibrillin	lumican	coll IV	TIMP-1	IGFBP	coll VI	TIMP-3	CTGF	hevin	fibulin	BM-40	TIMP-2	HSPG	fibronectin	MGP	C/DSPG	fibr-r	coll-I	coll-III	MMP
laminin		7	9	21	9	15	8	4	5	7	14	10	7	11	9	19	11	7	16	10	13
fibrillin	7		13	8	6	7	14	11	4	7	12	7	8	4	8	6	13	6	11	12	11
lumican	9	13		24	17	16	28	17	17	14	15	22	10	8	12	25	33	14	32	34	17
coll IV	21	8	24		17	22	22	13	11	14	28	25	12	22	16	27	26	12	34	25	26
TIMP-1	9	6	17	17		20	15	11	11	6	10	21	15	9	16	20	13	8	14	20	19
IGFBP	15	7	16	22	20		20	18	16	11	14	18	14	19	21	25	23	10	27	23	20
coll VI	8	14	28	22	15	20		13	17	19	16	20	11	11	19	19	28	12	31	36	27
TIMP-3	4	11	17	13	11	18	13		13	18	20	22	14	9	10	18	25	12	12	13	9
CTGF	5	4	17	11	11	16	17	13		8	10	18	7	7	19	22	12	12	18	13	11
hevin	7	7	14	14	6	11	19	18	8		15	18	13	8	8	23	27	10	14	11	8
fibulin	14	12	15	28	10	14	16	20	10	15		19	9	11	8	19	20	6	17	20	18
BM-40	10	7	22	25	21	18	20	22	18	18	19		14	11	24	21	24	16	25	32	19
TIMP-2	7	8	10	12	15	14	11	14	7	13	9	14		7	12	8	16	11	13	13	13
HSPG	11	4	8	22	9	19	11	9	7	8	11	11	7		8	14	10	6	11	10	10
fibronectin	9	8	12	16	16	21	19	10	19	8	8	24	12	8		14	14	11	24	21	15
MGP	19	6	25	27	20	25	19	18	22	23	19	21	8	14	14		32	14	25	20	13
C/DSPG	11	13	33	26	13	23	28	25	12	27	20	24	16	10	14	32		14	27	28	14
fibr-r	7	6	14	12	8	10	12	12	12	10	6	16	11	6	11	14	14		14	13	6
coll-I	16	11	32	34	14	27	31	12	18	14	17	25	13	11	24	25	27	14		42	21
coll-III	10	12	34	25	20	23	36	13	13	11	20	32	13	10	21	20	28	13	42		23
MMP	13	11	17	26	19	20	27	9	11	8	18	19	13	10	15	13	14	6	21	23	

#### 45 V Novel Genes Associated with Matrix Remodeling

Using coexpression analysis, we have identified 20 novel genes that show strong association with known matrix remodeling genes from a total of 41,419 assembled gene

sequences. The degree of association was measured by probability values and has a cutoff of p value less than 0.00001. This was followed by annotation and literature searches to insure that the genes that passed the probability test have strong association with known matrix-remodeling genes. This process was reiterated so that the initial 41,419 genes were reduced to the final 20 matrix-remodeling genes. Details of the coexpression patterns for the 20 novel matrix-remodeling genes are presented in Table 6.

Each of the 20 novel genes is coexpressed with at least two of the 21 known genes with a p-value of less than  $10^{-7}$ . The coexpression results are shown in Table 6.

The novel genes identified are listed in the table by their Incyte clone numbers (Clone), and the known genes their abbreviated names (Gene) as shown in Example IV.

Table 6. Coexpression of 20 novel genes with known matrix-remodeling genes. (- log p)

Gene	laminin	fibrillin	lumican	coll IV	TIMP-1	IGFBP	coll VI	TIMP-3	CTGF	hevin	fibulin	BM-40	TIMP-2	HSPG	fibronectin	MGP	C/DSP	fib-r	coll-I	coll-III	MMP
606132	8	7	2	6	4	7	7	2	4	4	4	3	3	4	4	3	2	2	5	3	10
627722	3	4	1	1	3	3	2	5	3	6	3	4	3	2	6	5	3	3	2	3	4
639644	6	7	11	10	3	4	7	3	14	6	6	9	6	2	9	8	5	6	9	7	6
1362659	6	5	6	7	6	9	10	9	8	8	7	6	8	6	7	9	9	7	10	5	5
1446685	6	6	11	13	4	7	8	5	7	5	10	9	5	9	5	9	8	6	8	10	7
1556751	3	7	7	8	8	9	9	8	7	6	5	5	7	8	4	10	11	3	7	6	8
1656953	6	8	6	2	5	7	8	5	6	9	3	7	4	3	4	10	8	7	4	4	5
1662318	9	3	6	10	7	9	5	5	8	8	6	8	5	9	6	8	6	4	7	7	9
1996726	3	4	7	7	6	5	8	3	10	2	2	3	2	2	9	3	6	6	8	11	6
2137155	3	2	6	3	4	2	2	4	6	4	2	9	4	2	8	4	4	4	5	2	5
2268890	9	13	7	9	8	11	8	9	5	5	8	7	8	5	8	8	11	3	11	7	11
2305981	3	2	4	6	3	4	3	5	5	6	7	5	2	2	2	7	6	4	3	2	2
2457612	3	3	3	5	2	4	4	2	8	4	5	5	2	2	7	8	6	6	5	4	8
2814981	6	3	5	7	4	6	7	2	2	5	5	5	3	6	5	4	6	1	6	4	7
3089150	4	6	11	8	5	10	13	9	14	10	11	10	7	6	8	11	16	11	9	7	5
3206667	8	5	10	9	7	5	6	4	9	4	7	8	4	4	7	13	12	4	8	8	6
3284695	7	6	7	14	8	7	6	14	8	18	12	9	10	8	6	18	10	5	13	6	6
3481610	3	2	4	4	3	6	4	6	6	7	4	5	1	5	5	7	5	3	3	2	2
3722004	6	4	8	10	13	9	7	13	8	9	11	12	11	5	10	9	12	3	7	7	6
3948614	11	8	6	17	8	13	12	5	5	11	12	7	11	13	4	7	7	4	14	11	10

## VI Novel Genes Associated with Matrix Remodeling

The 20 novel genes were identified from the data shown in Table 6 to be associated with matrix remodeling.

The nucleotide sequences comprising the consensus sequences of SEQ ID NOs:1-20 of the present invention were first identified from Incyte Clones 606132, 627722, 639644, 1362659, 1446685, 1556751, 1656953, 1662318, 1996726, 2137155, 2268890, 2305981, 2457612, 2814981, 3089150, 3206667, 3284695, 3481610, 3722004, and 3948614, respectively, and assembled according to Example III. BLAST and other motif searches were performed for SEQ ID NOs:1-20 according to Example VII. The sequences of SEQ ID NOs:1-20 were translated and sequence identity was sought with known sequences. Polypeptide sequences comprising the consensus sequences of SEQ ID NO:21, SEQ ID NO:22, and SEQ ID NO:23 of the present invention were encoded by SEQ ID NO:2,

SEQ ID NO:6, and SEQ ID NO:11, respectively. SEQ ID NOs:21-23 were analyzed using BLAST and other motif search tools as disclosed in Example VII.

SEQ ID NO:3 is 2987 residues in length and shows about 59% sequence identity from about nucleotide 2117 to about nucleotide 2914 with the cDNA encoding regulatory subunit of a human cAMP-dependent protein kinase, RIIbeta (WO 88/03164). SEQ ID NO:8 is 3017 nucleotides in length and shows about 70% to about 74% sequence identity from about nucleotide 1 to about nucleotide 1260 and about nucleotide 1925 to about nucleotide 1985 with human Hpast mRNA (g2529706), a gene associated with multiple endocrine neoplasia type 1. SEQ ID NO:9 is 1735 nucleotides in length and shows about 25% sequence identity from about nucleotide 5 to about nucleotide 1534 with a human neuronal cell adhesion molecule (WO 96/04396) important in the development of nervous system by promoting cell-cell adhesion. SEQ ID NO:14 is 2040 nucleotides in length and shows about 60% to 70% sequence identity from about nucleotide 1 to about nucleotide 1023 with a human mRNA for a serine protease (g1621243) specific for insulin-like growth factor-binding proteins. The amino acid sequence encoded by SEQ ID NO:14 from about nucleotide 3 to about nucleotide 1043 shows about 61% sequence identity with an osteoblast-like cell-derived protein (J09107980) useful for treatment and prevention of various diseases and as contraceptive. SEQ ID NO:15 is 2121 nucleotides in length and shows 60-80% sequence identity with a mouse gene, ADAMT-1 (g2809056), a member of the ADAM (the disintegrin and metalloproteinase) family. ADAMT-1 has been shown to contain the thrombospondin (TSP) type I motif; expression of ADAMT-1 is closely associated with inflammatory processes (Kuno et al (1997) Genomics 46:466-471). SEQ ID NO:16 is 2900 nucleotides in length and shows about 70% sequence identity with a mouse homeobox (Pmx) mRNA (g460124). Homeobox genes are expressed in very specific temporal and spatial pattern and function as transcriptional regulators of developmental processes (Kern et al. (1994) Genomics 19:334-340).

SEQ ID NO:21 is 551 amino acid residues long and shows about 37% sequence identity from about amino acid residue 10 to about amino acid residue 278 with PALM (g3219602), a human paralemin that is membrane-bound and expressed abundantly in brain and at intermediate levels in the kidney and in endocrine cells. In addition, the sequence encompassing residues 418 to 434 of SEQ ID NO:21 resembles one of the structural fingerprint regions of a seven trans-membrane receptor, LCR1, that is isolated from the human brain (Rimland et al. (1991) Mol Pharmacol 40:869-875). SEQ ID NO:21 also has one potential amidation site at L546; three potential N-glycosylation sites at N223, N229, and N408; one potential cAMP- and cGMP-dependent protein kinase phosphorylation site at S486; fifteen potential casein kinase II phosphorylation sites at S57, S100, T101, T116, S135, S253, T349, S370, T387, S426, T434, S489, S505, S520, and T526; one potential N-myristoylation site at G54; and nine potential protein kinase C phosphorylation sites at T15, S25, S57, S100, S123, S247,

S364, S370, and S505. SEQ ID NO:22 is 99 amino acid residues in length. The sequence of SEQ ID NO:22 from about amino acid residue 71 to about amino acid residue 81 resembles one of the fingerprint regions of the RH1 and RH2 opsins, a family of G protein coupled receptors that mediate vision (Zuker et al. (1985) Cell 40:851-858; Cowman et al. (1986) Cell 44:705-710). SEQ ID NO:22 also has one potential N-myristoylation site at G24, and two potential protein kinase C phosphorylation sites at S13 and S89. SEQ ID NO:23 is 493 amino acid residues in length and shows about 44% sequence identity from about amino acid residue 277 to about amino acid residue 487 with an angiopoietin-like factor from the human cornea, CDT6 (g2765527). Angiopoietin 1 and angiopoietin 2 function as a natural ligand and a natural inhibitor, respectively, for TIE2, a receptor critical in angiogenesis during embryonic development, tumor growth, and tumor metastasis. The sequences encompassing amino acid residues 305 to 343, 346 to 355, 365 to 402, 411 to 424, and 428 to 458 of SEQ ID NO:23 resemble the carboxy-terminal domain signatures of fibrinogen beta and gamma chains from BLOCKS analysis. SEQ ID NO:23 also exhibits one potential signal peptide region encompassing amino acid residues M1 to G22 when analyzed using a HMM-based signal peptide analysis tool. In addition, SEQ ID NO:23 shows two potential N-glycosylation sites at N164 and N192; one potential cAMP- and cGMP-dependent protein kinase phosphorylation sites at S127, six potential casein kinase II phosphorylation sites at S34, S209, T238, S266, T368, and T417; four potential N-myristoylation sites at G12, G18, G22, and G29; eight potential protein kinase C phosphorylation sites at S34, S209, T268, T299, T335, S373, S383, and S477; and three potential tyrosine kinase phosphorylation sites at Y183, Y392, and Y467.

## **VII Homology Searching for Matrix-Remodeling Renes and the Proteins Encoded by the Genes**

Polynucleotide sequences, SEQ ID NOs:1-20, and polypeptide sequences, SEQ ID NOs: 21-23, were queried against databases derived from sources such as GenBank and SwissProt. These databases, which contain previously identified and annotated sequences, were searched for regions of similarity using Basic Local Alignment Search Tool (BLAST; Altschul (1990) supra) and Smith-Waterman alignment (Smith et al. (1992) Protein Engineering 5:35-51). BLAST searched for matches and reported only those that satisfied the probability thresholds of  $10^{-25}$  or less for nucleotide sequences and  $10^{-8}$  or less for polypeptide sequences.

The polypeptide sequences were also analyzed for known motif patterns using MOTIFS, SPSCAN, BLIMPS, and Hidden Markov Model (HMM)-based protocols. MOTIFS (Genetics Computer Group, Madison WI) searches polypeptide sequences for patterns that match those defined in the Prosite Dictionary of Protein Sites and Patterns (Bairoch et al. supra), and displays the patterns found and their corresponding literature abstracts. SPSCAN (Genetics Computer Group) searches for

potential signal peptide sequences using a weighted matrix method (Nielsen et al. (1997) Prot Eng 10:1-6). Hits with a score of 5 or greater were considered. BLIMPS uses a weighted matrix analysis algorithm to search for sequence similarity between the polypeptide sequences and those contained in BLOCKS, a database consisting of short amino acid segments, or blocks, of 3-60 amino acids in length, compiled from the PROSITE database (Henikoff et al. supra; Bairoch et al. supra), and those in PRINTS, a protein fingerprint database based on non-redundant sequences obtained from sources such as SwissProt, GenBank, PIR, and NRL-3D (Attwood et al. (1997) J Chem Inf Comput Sci 37:417-424). For the purposes of the present invention, the BLIMPS searches reported matches with a cutoff score of 1000 or greater and a cutoff probability value of  $1.0 \times 10^{-3}$ . HMM-based protocols were based on a probabilistic approach and searched for consensus primary structures of gene families in the protein sequences (Eddy, supra; Sonnhammer, supra). More than 500 known protein families with cutoff scores ranging from 10 to 50 bits were selected for use in this invention.

#### VIII Labeling and Use of Individual Hybridization Probes

Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250  $\mu$ Ci of [ $\gamma$ - $^{32}$ P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (NEN Life Science Products, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine resin column (Amersham Pharmacia Biotech). An aliquot containing  $10^7$  counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (NEN Life Science Products).

The DNA from each digest is fractionated on a 0.7 percent agarose gel and transferred to nylon membranes (NYTRAN PLUS, Schleicher & Schuell, Durham NH). Hybridization is carried out under the following conditions: 5x SCC/0.1% SDS at 60° C for about 6 hours, subsequent washes are performed at higher stringency with buffers, such as 1x SCC/0.1% SDS at 45° C, then 0.1xSCC. After XOMAT AR film (Eastman Kodak, Rochester NY) is exposed to the blots for several hours, hybridization patterns are compared.

#### IX Production of Specific Antibodies

SEQ ID NO:20, 21, or 23 substantially purified using polyacrylamide gel electrophoresis (Harrington (1990) Methods Enzymol 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the amino acid sequence is analyzed using LASERGENE software (DNASTAR, Madison WI) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for

selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. Typically, oligopeptides 15 residues in length are synthesized using an ABI 431A peptide synthesizer (PE Biosystems) using Fmoc-chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester to increase

- 5 immunogenicity. Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide activity by, for example, binding the peptide to plastic, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radioiodinated goat anti-rabbit IgG.

What is claimed is:

1. A substantially purified polynucleotide comprising a gene that is coexpressed with one or  
5 more known matrix-remodeling genes in a plurality of biological samples, wherein each known  
matrix-remodeling gene is selected from the group consisting of osteonectin, chondroitin/dermatan  
sulfate proteoglycans, collagen I, II, II, and IV, connective tissue growth factor, fibrillin, fibronectins,  
fibronectin receptor, fibulin 1, heparan sulfate proteoglycan, extracellular matrix protein, insulin-like  
growth factor 1, insulin-like growth factor binding protein, laminin, lumican, matrix Gla protein,  
10 matrix metalloproteases, and tissue inhibitors of matrix metalloproteinase 1, 2, and 3.

2. The polynucleotide of claim 1, comprising a polynucleotide sequence selected from the  
group consisting of:

- (a) a polynucleotide sequence selected from the group consisting of SEQ ID NOs:1- 20;
- (b) a polynucleotide sequence which encodes the polypeptide sequence of SEQ ID NO: 21,  
15 22, or 23;
- (c) a polynucleotide sequence having at least 70% identity to the polynucleotide sequence of  
(a) or (b);
- (d) a polynucleotide sequence comprising at least 18 sequential nucleotides of the  
polynucleotide sequence of (a), (b), or (c);
- 20 (e) a polynucleotide which hybridizes under stringent conditions to the polynucleotide of (a),  
(b), (c), or (d); and
- (f) a polynucleotide sequence which is complementary to the polynucleotide sequence of (a),  
(b), (c), (d), or (e).

3. A substantially purified polypeptide comprising the gene product of a gene that is  
25 coexpressed with one or more known matrix-remodeling genes in a plurality of biological samples,  
wherein each known matrix-remodeling gene is selected from the group consisting of osteonectin,  
chondroitin/dermatan sulfate proteoglycans, collagen I, II, II, and IV, connective tissue growth factor,  
fibrillin, fibronectins, fibronectin receptor, fibulin 1, heparan sulfate proteoglycans, extracellular  
matrix protein, insulin-like growth factor 1, insulin-like growth factor binding protein, laminin,  
30 lumican, matrix Gla protein, matrix metalloproteases, and tissue inhibitors of matrix metalloproteinase  
1, 2, and 3.

4. The polypeptide of claim 3, comprising a polypeptide sequence selected from the group  
consisting of:

- (a) the polypeptide sequence of SEQ ID NO:21, 22, or 23;
- 35 (b) a polypeptide sequence having at least 85% identity to the polypeptide sequence of (a);  
and



(c) a polypeptide sequence comprising at least 6 sequential amino acids of the polypeptide sequence of (a) or (b).

5. An expression vector comprising the polynucleotide of claim 2.

6. A host cell comprising the expression vector of claim 5.

5 7. A pharmaceutical composition comprising the polynucleotide of claim 2 or the polypeptide of claim 3 in conjunction with a suitable pharmaceutical carrier.

8. An antibody which specifically binds to the polypeptide of claim 4.

9. A method for diagnosing a disease or condition associated with the altered expression of a gene that is coexpressed with one or more known matrix-remodeling genes, wherein each known  
10 matrix-remodeling gene is selected from the group consisting of osteonectin, chondroitin/dermatan sulfate proteoglycans, collagen I, II, III, and IV, connective tissue growth factor, fibrillin, fibronectins, fibronectin receptor, fibulin 1, heparan sulfate proteoglycans, extracellular matrix protein, insulin-like growth factor 1, insulin-like growth factor binding protein, laminin, lumican, matrix Gla protein, matrix metalloproteinases, and tissue inhibitors of matrix metalloproteinase 1, 2, and 3, the method  
15 comprising the steps of:

(a) providing a sample comprising one or more of said coexpressed genes;

(b) hybridizing the polynucleotide of claim 2(F) to said coexpressed genes under conditions effective to form one or more hybridization complexes; and

(c) detecting the hybridization complexes, wherein the altered level of hybridization  
20 complexes compared with the level of hybridization complexes of a nondiseased sample correlates with the presence of the disease or condition.

10. A method for treating or preventing a disease associated with the altered expression of a gene that is coexpressed with one or more known matrix-remodeling genes in a subject in need, wherein each known matrix-remodeling gene is selected from the group consisting of osteonectin,  
25 chondroitin/dermatan sulfate proteoglycans, collagen I, II, III, and IV, connective tissue growth factor, fibrillin, fibronectins, fibronectin receptor, fibulin 1, heparan sulfate proteoglycans, extracellular matrix protein, insulin-like growth factor 1, insulin-like growth factor binding protein, laminin, lumican, matrix Gla protein, matrix metalloproteinases, and tissue inhibitors of matrix metalloproteinase 1, 2, and 3, the method comprising the step of administering to said subject in need the pharmaceutical  
30 composition of claim 7 in an amount effective for treating or preventing said disease.

11. A method for treating or preventing a disease associated with the altered expression of a gene that is coexpressed with one or more known matrix-remodeling genes in a subject in need, wherein each known matrix-remodeling gene is selected from the group consisting of osteonectin, chondroitin/dermatan sulfate proteoglycans, collagen I, II, III, and IV, connective tissue growth factor,

fibrillin, fibronectins, fibronectin receptor, fibulin 1, heparan sulfate proteoglycans, extracellular matrix protein, insulin-like growth factor 1, insulin-like growth factor binding protein, laminin, lumican, matrix Gla protein, matrix metalloproteases, and tissue inhibitors of matrix metalloproteinase 1, 2, and 3, the method comprising the step of administering to said subject in need the antibody of  
5 claim 8 in an amount effective for treating or preventing said disease.

12. A method for treating or preventing a disease associated with the altered expression of a gene that is coexpressed with one or more known matrix-remodeling genes in a subject in need, wherein each known matrix-remodeling gene is selected from the group consisting of osteonectin, chondroitin/dermatan sulfate proteoglycans, collagen I, II, III, and IV, connective tissue growth factor,  
10 fibrillin, fibronectins, fibronectin receptor, fibulin 1, heparan sulfate proteoglycans, extracellular matrix protein, insulin-like growth factor 1, insulin-like growth factor binding protein, laminin, lumican, matrix Gla protein, matrix metalloproteases, and tissue inhibitors of matrix metalloproteinase 1, 2, and 3, the method comprising the step of administering to said subject in need the polynucleotide sequence of claim 2(F) in an amount effective for treating or preventing said disease.

## SEQUENCE LISTING

<110> INCYTE PHARMACEUTICALS, INC.  
WALKER, Michael G.  
VOLKMUTH, Wayne  
KLINGLER, Tod M.

<120> MATRIX-REMODELING GENES

<130> PB-0004 PCT

<140> To Be Assigned  
<141> Herewith

<150> 09/169,289

<151> 1998-10-09

<160> 23

<170> PERL Program

<210> 1  
<211> 1447  
<212> DNA  
<213> Homo sapiens

<220>  
<221> unsure  
<222> 1380  
<223> a or g or c or t, unknown, or other

<220>  
<221> misc\_feature  
<223> Incyte ID No.: 606132CB1

<400> 1  
cctggaacca gaaggagacc tacctgcaca tcatgaagaa cgaggaggag gtggtgatct 60  
tgttcgcgca ggtgggcgac cgcagcatca tgcaaagcca gagcctgatg ctggagctgc 120  
gagagcagga ccaggtgtgg gtacgcctct acaagggcga acgtgagaac gccatcttca 180  
gcgaggagct ggacacctac atcaccttca gtggctacct ggtcaagcac gccaccgagc 240  
cctagctggc cggccacctc ctttcctctc gccaccttcc acccctgcgc tgtgctgacc 300  
ccaccgcctc ttccccgata cctggactcc gactccctgg ctttggcatt cagtggagacg 360  
ccctgcacac acagaaagcc aaagcgatcg gtgctcccag atcccgcagc ctctggagag 420  
agctgacggc agatgaaatc accagggcgg ggcacccgcg agaaccctct gggaccttcc 480  
gcgccctct ctgcacacat cctcaagtga ccccgacgg cgagacgcgg gtggcggcag 540  
ggcgtcccag ggtgcggcac cgcggctcca gtccttgaa ataattaggc aaattctaaa 600  
ggtctcaaaa ggagcaaaagt aaaccgtgga ggacaaagaa aagggttggt atttttgtct 660  
ttccagccag cctgctggct cccaagagag aggcctttc agttgagact ctgcttaaga 720  
gaagatccaa agttaagct ctggggtcag gggaggggccc gggggcagga aactacctct 780  
ggcttaattc ttttaagcca cgtaggaact ttcttgagg ataggtggac cctgacatcc 840  
ctgtggcctt gcccagggc tctgctggc tttctgagtc acagctgcga ggtgatggg 900  
gctggggccc caggcgtcag ctcccagagg gacagctgag cccctgcct tggtccagg 960  
ttggtagaag cagccgaagg gctcctgaca gtggccagg acccctgggt ccccaggcc 1020  
tgacagatgt tctatgagg gcagagctcc tggatcatcc atgtgtggct ctgctccacc 1080  
cctgtgccac cccagagccc tggggggtgg tctccatgcc tgccaccctg gcacggctt 1140  
tctgtgccgc ctcccacaca aatcagcccc agaaggcccc ggggccttgg cttctgtttt 1200  
ttataaaaca cctcaagcag cactgcagtc tcccatctcc tcgtgggcta agcatcaccg 1260  
cttccacgtg tggtgtgtg gttggcagca aggtgatcc agacccttc tgccccact 1320  
gcgctcatcc aggcctctga ccagtagcct gagaggggct tttctaggc ttcagagcan 1380  
gggagagctg gacggggtag acagtccgct tgtctgttct aagctctgtg agctcagct 1440  
gagacaa 1447

<210> 2  
<211> 2481  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc feature  
<223> Incyte ID No.: 627722CB1

<400> 2  
ctagcaagca ggtaaacgag ctttgtacaa acacacacag accaacacat ccggggatgg 60  
ctgtgtgttg ctagagcaga ggctgattaa acactcagtg tgttggtctt ctgtgccact 120  
cctggaaaat aatgaattgg gtaaggaaca gttaataaga aaatgtgcct tgctaactgt 180  
gcacattaca acaaagagct ggcagctcct gaaggaaaag ggcttgtgcc gctgccgttc 240  
aaacttgtca gtcaactcat gccagcagcc tcagcgtctg cctccccagc acaccctcat 300  
tacatgtgtc tgtctggcct gatctgtgca tctgtcggga gacgctcctg acaagtctgg 360  
aatcttctta tttctccact ggtgcaaaga gcggatttct ccctgcttct cttctgtcac 420  
ccccgtcctt ctcccccagg aggtctccttg atttatggta gctttggact tgcttccccg 480  
tctgactgtc cttgacttct agaatggaag aagctgagct ggtgaaggga agactccagg 540  
ccatcacaga taaaagaaaa atacaggaag aaatctcaca gaagcgtctg aaaatagagg 600  
aagacaaact aaagaccag catttgaaga aaaaggcctt gagggagaaa tggcttctag 660  
atggaatcag cagcggaaaa gaacaggaag agatgaagaa gcaaaatcaa caagaccagc 720  
accagatcca ggttctagaa caaagtatcc tcaggcttga gaaagagatc caagatcttg 780  
aaaaagctga actgcaaatc tcaacgaagg aagaggccat tttaaagaaa cttaaagtaa 840  
ttgagcggac aacagaagac attataagat ctgtgaaagt ggaaagagaa gaaagagcag 900  
aagagtcaat tgaggacatc tatgctaata tccctgacct tccaaagtcg tacatacctt 960  
ctaggttaag gaaggagata aatgaagaaa aagaagatga tgaacaaaat aggaaagctt 1020  
tatatgccat ggaaattaaa gttgaaaaag acttgaagac tggagaaagt acagttctgt 1080  
cttcaatacc tctgccatca gatgacttta aaggtacagg aataaaaagt tatgatgatg 1140  
ggcaaaagtc agtgtatgca gtaagttcta atcacagtgc agcatacaat ggcaccgatg 1200  
gcctggcacc agttgaagta gaggaacttc taagacaagc ctgagagaga aactctaaat 1260  
ccccacaga gtatcatgag cctgtatatg ccaatccctt ttacaggcctt acaaccccac 1320  
agagagaaac ggtgaccctt ggaccaaact ttcaagaaag gataaagatt aaaactaatg 1380  
gactgggtat tgggtgtaaat gaatccatac acaatatggg caatgggtctt tcagaggaaa 1440  
ggggaaacaa cttcaatcac atcagtcaca ttccggcagt gcctcatccc cgatcagtga 1500  
ttcaacaagc agaagagaag cttcacaccc cgcaaaaaag gctaatagact ccttgggaaag 1560  
aatcgaatgt catgcaggac aaagatgcac cctctccaaa gccaaaggctg agccccagag 1620  
agacaatatt tgggaaatct gaacaccaga attcttcacc cacttgtcag gaggacgagg 1680  
aagatgtcag atataatata gtatcattccc tgctccaga cataaatgat acagaaccgg 1740  
tgacaatgat tttcatgggg tatcagcagg cagaagacag tgaagaagat aagaagtttc 1800  
tgacaggata tgatgggatc atccatgctg agctgggtgt gattgatgat gaggaggagg 1860  
aggatgaagg agaagcagag aaaccgtcct accaccccat agctcccat agtcagggtgt 1920  
accagccagc caaaccaaca ccacttccta gaaaagatc agaagctagt cctcatgaaa 1980  
acacaaatca taaatcccc caaaaaatt ccatatctct gaaagagcaa gaagaaagct 2040  
taggcagccc tgtccacatc tcccatttg atgctcagac aactggagat gggactgagg 2100  
atccatcctt aacagcttta aggatgagaa tggcaaaagt gggaaaaaag gtgatctaag 2160  
agttgtacca cctatataaa catcctttga agaagaaact aagaagcatt tgcaaatctc 2220  
tcttctggat atttgttta tttttctga agtccaaaaa attatcatta cagtgtacca 2280  
tattaagcca tgtgaataag tagtagtcat tatttgtgaa aaattcccaa aaagctgggg 2340  
aaaacaaatg tgtaactttt ccagttactt gacacgattc agtgggggaa aaccagcatt 2400  
ttttattcta ttgataccaa agcatttcta ataagagctt gttaaattta agaataaagt 2460  
tatttaaaat aaaaaaaaaa a 2481

<210> 3  
<211> 2987  
<212> DNA  
<213> Homo sapiens

<220>  
<221> unsure

&lt;222&gt; 2955

&lt;223&gt; a or g or c or t, unknown, or other

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No.: 639644CB1

&lt;400&gt; 3

```
agaaaaaaag aaaaaagaaa aaaactaagg cagcagctct taataaataa cacctggagc 60
agaatcggta aactgctttc acgttggcct ttgcagaagt ggcaatgcat tgaggataca 120
tctggcaagc ttcgaattca caagtgtaaa ggaccagtg acctgctcac agtccggcag 180
agcacgcgga acctctacgc tcgcggtctc catgacaaag acaaagagtg cagttgtagg 240
gagctctggt accgtgccag cagaagccaa agaaagagtc aacggcaatt cttgagaaac 300
caggggactc caaagtacaa gccagattt gtccatactc ggcagacacg ttccttgctc 360
gtcgaatttg aaggtgaaat atatgacata aatctggaag aagaagaaga attgcaagtg 420
ttgcaaccaa gaaacattgc taagcgtcat gatgaaggcc acaaggggcc aagagatctc 480
caggttcca gtggtggcaa caggggcagg atgctggcag atagcagcaa cgccgtgggc 540
ccacctacca ctgtccgagt gacacacaag tgttttattc ttccaatga ctctatccat 600
tgtgagagag aactgtacca atcgccaga gcgtggaagg accataaggc atacattgac 660
aaagagattg aagctctgca agataaaaatt aagaatttaa gagaagtgag aggacatctg 720
aagagaagga agcctgagga atgtagctgc agtaaacaaa gctattacaa taaagagaaa 780
ggtgtaaaaa agcaagagaa attaaagagc catcttcacc cattcaagga ggctgctcag 840
gaagtagata gcaaactgca acttttcaag gagaacaacc gtaggaggaa gaaggagagg 900
aaggagaaga gacggcagag gaagggggaa gagtgcagcc tgcctggcct cacttgcttc 960
acgcatgaca acaaccactg gcagacagcc ccgttctgga acctgggac tttctgtgct 1020
tgacagagtt ctaacaataa cacctactgg tgtttgcgta cagttaatga gacgcataat 1080
tttcttttct gtgagtttgc tactggcttt ttggagtatt ttgatatgaa tacagatcct 1140
tatcagctca caaatcagat gcacacgcta gaacgaggca ttttgaatca gctacacgta 1200
caactaatgg agctcagaag ctgtcaagga tataagcagt gcaaccaag acctaagaat 1260
cttgatgttg gaaataaaga tggaggaagc tatgacctac acagaggaca gttatgggat 1320
ggatgggaag gttaatcagc ccgctctcac tgcagacatc aactggcaag gcctagagga 1380
gctacacagt gtgaatgaaa acatctatga gtacagacaa aactacagac ttagtctggt 1440
ggactggact aattacttga aggattttag tagagtattt gcaactgctga agagtacta 1500
tgagcaaaat aaaacaaata agactcaaac tgcctcaagt gacgggttct tggttgtctc 1560
tgctgagcac gctgtgtcaa tggagatggc ctctgctgac tcagatgaag acccaaggca 1620
taaggttggg aaaacacctc atttgacctt gccagctgac cttcaaaacc tgcatattgaa 1680
ccgaccaaca ttaagtccag agagtaaact tgaatggaat aacgacattc cagaagttaa 1740
tcatttgaat tctgaacact ggagaaaaac cgaaaaatgg acggggcatg aagagactaa 1800
tcatctgtaa accgatttca gtggcgatgg catgacagag cttagagctg ggcccagccc 1860
caggctgcag ccatttcgca ggcaccgaa agaactccc cagtatggtg gtccctggaaa 1920
ggacattttt gaagatcaac tatatcttcc tgtgcattcc gatggaattt cagttcatca 1980
gatgttcacc atggccaccg cagaacaccg aagtaattcc agcatagcgg ggaagatggt 2040
gaccaaggtg gagaagaatc acgaaaagga gaagtcacag cacctagaag gcagcgcttc 2100
ctcttcactc tcctctgatt agatgaaact gttaccttac cctaaacaca gtatttcttt 2160
ttaacttttt tatttgtaaa ctaataaagg taatcacagc caccaacatt ccaagctacc 2220
ctgggtacct ttgtgcagta gaagctagt agcatgtgag caagcgggtg gcacacggag 2280
actcatcggt ataatttact atctgccaag agtagaaaga aaggctgggg atatttgggt 2340
tggcttggtt ttgatttttt gcttggttgt ttgttttgta ctaaaacagt attatctttt 2400
gaatatcgta gggacataag tatatacatg ttatccaatc aagatggcta gaatggtgcc 2460
tttctgagtg tctaaaactt gacaccctg gtaaatcttt caacacactt ccactgcctg 2520
cgtaatgaag ttttgattca tttttaacca ctggaatttt tcaatgccgt cattttcagt 2580
tagatgattt tgcactttga gattaaaatg ccatgtctat ttgattagtc ttattttttt 2640
atttttacag gcttatcagt ctcactgttg gctgtcattg tgacaaagtc aaataaaccc 2700
ccaaggacga cacacagtat ggatcacata ttgtttgaca ttaagctttt gccagaaaat 2760
gttgcatgtg ttttacctcg acttgctaaa atcgattagc agaaaggcat ggctaataat 2820
gttggtggtg aaaaataata aataagtaaa caaaaaaaa aaaaaaaa aaaaaaaa 2880
aaaaaaaaa aaaaaaaa aaaaagcaaa aaaagctgcc gccacagtta gatgaagaag 2940
catgaggatc cgagnnggtc gcctctttga gtggtgaggg agtcgcg 2987
```

&lt;210&gt; 4

&lt;211&gt; 2915

<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No.: 1362659CB1

<400> 4  
gaggcaagaa ttccggcacga gggacatddd gccaaacttaa acgagaaaaa gacccccccgc 60  
acccggcaca ctcccccttc ctccagcccc gcttcagcca catgctccag ctgctgcccc 120  
gtaaagccct gtgccttttt ttcccctgaa tactgccaa agcatccct tcccactgc 180  
ctctcaggag ttggggactt tgctaggaga ttttttaagt gttccttact gggacaacgt 240  
ggagccacgt ttgcaggagc tccatttcta tccctgctgg tgttgacttc tgtgtagggg 300  
ccagttcatg tccctgactc tcacctccca ttagataaat gaagcccacc cccctttcta 360  
gagtgatgag agtcaagaag aggggatgta tgaacggcca aattcccatg tgagaggaag 420  
atgacctgat ccacctagcc tttcttctg gatctgtcct cctcacccc tttcacctga 480  
gctgtccaca gtaggaaaca taaagaaaca atgtccccta catatcccca tgactacata 540  
atccatcatc gtaggaaata ggaaagcaaa tttgattdtg gttttgtaa acgtacatgc 600  
ttcaataatt ctttttttgt gtcttaataa ctcatagggg aaaaaaacag ctccaccaag 660  
gtgttaggtt tcacatatat attcatcaac tatttttaga gatttaattc tatcaaatct 720  
tgtattacct cagatcattt taaatagcaa gccaatacg agctttgaag gctattttac 780  
cattcctgtt cacaagaggt tctcatgggt cctgacaggt tacccttgag ggcttggtc 840  
tactttttta agtcaatgg tttttttct tgtgttctag tttccataat aggagagaaa 900  
atatagaaat atatgcaaaa attatagttt tcttttagatc agaaactgat atttttgggt 960  
cagccatag tattttgttt aaaggattta aaataaagtg ccgtcatgta gccctgtgga 1020  
agggagcaca taaccagctg tttggcatga caggtgactt agtatatttg taattggttt 1080  
taaaaccaat acaccatact ttctttctgc aaacagccat ctttatactt agggaagaaa 1140  
aattgttggg ttctagactt ttttaataa aattttgttg atatggaatt aggttaagttt 1200  
aagtgtctat gtgcatatgt tttttatata agttttttct attcagtttc actgatccaa 1260  
ctggcagtg gtaaatatgg cataagttta taacactttt ccccaaaatg gtgctttgga 1320  
tttgaaaagg gtctgatggg gagaaggaga acgtatcatc ctgacttctt ctcttaataa 1380  
acctagaaaa acgggtagta aactgtggat agtcaggaaa acaccagca agggacacag 1440  
ctgtcaggaa atgaatcttc ccccaacccc ccacctgca gatggataga cagaatcttt 1500  
cctgactagt cattagatc aggggctct gttggatttg tgtttcttga agaataagctg 1560  
gcagagtgg ataaaagaca cgaatatctc tccgtctata aggatactct gatttgggtt 1620  
ttgcattttt catggttttt atttctgtt ccccctggag ttttccatta gtgagtttt 1680  
gtgcaaggat cttattttgt atgccttccc tcccctagaa agattttgtg caatatatta 1740  
aatggggaca gaattctaaa tggataaaac aatggctggg tctagccctg agtgacagtc 1800  
ttaaggctag atccttccca tagtatcatc tgtcctctgg aatgactctc ctgtccctaa 1860  
aggggttaag agagagatca cctagaaatc cctctggaca cttgtgggtt ctttagggtt 1920  
tgagtttctt cttcccttg agcttcagag aggaagttg gcatgggtta atctgaatgg 1980  
ttacctact gctgaaaacc cagaggggag tggcacactc gcttggttg aaaagcctct 2040  
aaatgcatcc cttcctttct ttcctgcttc ctttgctta caattgaagc agcccgtgg 2100  
accatcacag tatgcagaga cttcctcacc tttcatatct agggaccacc cccgatgcat 2160  
tggtaggggt gggcacttat aaatgcctgc tattgttaag ccattccagc ctcttctct 2220  
gaatagacca gacgcccttt cacttagttc agtgccagtc cttttgcctt cccaacctg 2280  
ctgttaggcc tgcgtttccc tttgctctg attaggagag atggaaggag atgagctccc 2340  
ataactgaat tggccttttg ttcatgtttt ctcccataat gtatatatgc catatgtgaa 2400  
tatgccatat atatgtgcca acaaatctat ctacgttgtt cttttcaaat tagcacgcag 2460  
ataggaattt tgagtttctt cttcttttag taactagtat aacaagcact ggtatttttg 2520  
tacaataaag aaaaacaaaa gattgactat tgtggtctgc atgacataaa caaacaattg 2580  
gtgatatcaa agcaacgtat accccagtc agtgtgtgtt gccataattt gcaattcagc 2640  
ttaacagtgc acccaatcta tatttgcatt ttgatattat ttaagctcta tgtacaaggt 2700  
tttgcatgta tttatatggg tcttagggaa aaaaaatgct ataaactgca aatctgaaat 2760  
tcaaatgtgt tgttccactg agaccagaag aagaagagga gttttaaaag ggataatttg 2820  
ttggagctaa taaagctttt tgctgatgaa cagaaaccaa tactgctgtg cactgagaat 2880  
aaaaactcat gccacttgt aaaaaaaaaa aaagg 2915

<210> 5

<211> 1826  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No.: 1446685CB1

<400> 5  
gaaagccgca gcctcagtcg cgccgcccgc cgctgcgtcc gccagcgcc agctccgcgt 60  
cccgaccggc ccgcccgcgc ctgcgcccgc ccatggccac ctccccgcag aagtcgcctt 120  
ctgtcpccaa gtctcccact cccaagtgcg ccccgctccg caagaaagat gattccttct 180  
tggggaaact cggaggggacc ctggcccggg ggaagaaagc caaggagggtg tccgagctgc 240  
aggaggaggg aatgaacgcc atcaacctgc cctcagccc aattcccttt gagctggacc 300  
ccgaggacac gatgctggag gagaatgagg tgcgaacaat ggtggatcca aactcacgca 360  
gtgaccccaa gcttcaagaa ctgatgaagg tattaattga ctggattaat gatgtgttg 420  
ttggagaaag aatcattgtg aaagacctag ctgaagattt gtatgatgga caagtccctgc 480  
agaagctttt cgagaaactg gagagtgaag agctaaatgt ggctgagggtc acccagtcag 540  
agattgtctc gaagcaaaaa ctgcagactg tcctggagaa gatcaatgaa accctgaaac 600  
ttcctcccag gagcatcaag tggaatgtgg attctgttca tgccaagagc ctggtggcca 660  
tcttacacct gctcgttgcg ctgtctcagt atttccgcgc accaattcga ctcccagacc 720  
atgtttccat ccaagtgtgt gtggtccaga aacgagaagg aatcctccag tctcggcaaa 780  
tccaagagga aataactggt aacacagagg ctctttccgg gaggcatgaa cgtgatgcct 840  
ttgacacctt gttcgacctg gcccagaca agctgaatgt ggtgaaaaag acactcatca 900  
ctttcgtgaa caagcacctg aataaaactg acctggaagt cacagaactg gaaacccagt 960  
ttgcagatgg ggtgtacctg gtgctgctca tggggctcct ggagggtac tttgtgcccc 1020  
tgcacagctt ctctctgacc ccgacagct ttgaacagaa ggtcttgaat gtctcctttg 1080  
cctttgagct catgcaagat ggagggttgg aaaagccaaa accgcggcca gaagacatag 1140  
tcaactgtga cctgaaatct acactacgag tgttgtacaa cctcttcacc aagtaccgta 1200  
acgtggagtg aggggctgcc ctgggcccac cactgcccac gagttcttgc tgttggcgta 1260  
ctggaccctc ctccgaactg ccttaccctg cttattcctg tctcttgac tgtgctctcc 1320  
cacaagtcca gctgcaaccc agagatagtg gaaactgaaa ttaggaagga aatcatcaat 1380  
aactcagtg gctgacccat cctcccagg cgctggggac caacctagca atgaagggtg 1440  
ggaagggtgt tcccttcccg gtgccaggtc cagatttccc tccatgattt gggaaccagg 1500  
ttaggcaaaa gactccccc aagatgaaaa taaagatcct agttaccatt caaaggatgc 1560  
taactgtgtg tcaggcccca cactaagtgc tctgctctga tatactcaag gccattaatc 1620  
ttcaggactc ccattgacgt aggtgtttca ttcccctttt acagatgagg aaactaaggc 1680  
ttggagggtta aatgacttgc cagaagttgg aatttttttc ctctttgaac ataacctctc 1740  
ccttctccct aaaggtaacc actattctga gtccaatcat caagggtttg cttttctttt 1800  
tagctaagta tgcattcctc aatagt 1826

<210> 6  
<211> 1439  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No.: 1556751CB1

<400> 6  
gagtatccct tgtttaatca cttttgtggt taaaagagac ctttgggtca gtctgcctca 60  
ttccttgaag agtttagccc tggctcactt ttactctat ttcttctcct gtctcaagaa 120  
agaagaaaaa aagagacaaa ttaccagaa acccctccct tccccacatg gaggccttgg 180  
caaatgttaa ttttctaga aaatccttca gacctgaaga cgcaggaaaa gaatctggct 240  
ctcagggttg cttctgcgtc cccgcccga ggcccagac tatggtcaca gggcgcctct 300  
gttctctccc gggactccag aatttctctc ctcaaaggaa agaaaacagg gcatgcgctt 360  
gttgcaaaaa cgcaggccg gctcccaaaa accccatgtg tgtacgatta aaagttggcc 420  
gtccccaggc ctcccagcgc aaacttaaa agacagggtt ttgctgaaaa ccaaactagg 480  
gccagctggg ctttttaaca acctagagac tttccggagc tgcctggaac agagcctgcg 540  
ggaaaacggg cttgccagag aactcacag tttccttcat ggcctgtttt ggtcccctaa 600  
gaatctccac atcattgtct ttcttgtgcc ttttcttgg tgagcaacag aaagggaagg 660

gttccaagcc tctaaaaatg tgctttgtga tcaggagtgc gctccaaacc aaatacgcgc 720  
gctgcccttt cgaggccagt gagctcagcc tccaaggctt taaagccaca tttcagcaag 780  
agaaagcgct gagagctcgc aggttcatta aagaaggcaa agcactgggt tctctcctta 840  
gaaaagtagg tttcttggtc tgatgtagac tggcttgctt tgatttttag tgaagggaat 900  
gtacgtaaaa caaaataggc cttggctggt caaaggagac aagcaggatg gatggatgga 960  
tggatggatg gatgtatgga tgaatagata gatgggtgtt gcatgtaaat tgcagagaaa 1020  
acaaaaccaa agctgattgg aaacaattaa ttgtgggtgt ctgaggggga aggtcgcagc 1080  
tttgggcagc tttgagaagc ggtacaagag ttctgtgcct gtgtgtccag ccctggagcc 1140  
agccagtga tttattttta gctcttagaa gcaactcctt ggcccaggaa tgcgtgacct 1200  
ctgagatggg tccacgcac tctctacact tcctctctc cgtgggatac tggactcgtg 1260  
cctctgcgcc cattctcttc tcacgcata ccatgagctt taatttctact tctctgacac 1320  
ggtacgtcga taaagccagt attacactta aatgaagtat tctttttgtt aatcgttttt 1380  
tttagaaggt aaacaaattt aataaagcta ccaataatga gaaaaaaaaa aaaaaaaaaa 1439

&lt;210&gt; 7

&lt;211&gt; 3047

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No.: 1656953CB1

&lt;400&gt; 7

cgagacagag gaaatgtgtc tccctccaag gccccaaagc ctgagagaaa ggggtgtttct 60  
ggttttgcct tagcaatgca tcggtctctg aggtgacact ctggagcggg tgaagggcca 120  
caaggtgcag ggttaatact cttgccagtt ttgaaatata gatgctatgg ttcagattgt 180  
ttttaataga aaactaaagg ggcaggggaa gtgaaaggaa agatggaggt tttgtgcggc 240  
tcgatggggc atttggaaact tctttttaaa gtcatctcat ggtctccagt tttcagttgg 300  
aactctgggt ttttaactt aagggagaca aaggctgtgt ccatttgga aaacttcctt 360  
ggccacgaga ctctaggtga tgtgtgaagc tgggcagtct gtggtgtgga gacgagccat 420  
ctgtctggcc attcagagga ttctaagac atggctggat gcgctgctga ccaacatcag 480  
cacttaataa atgcaaatg caacatttct ccctctgggc cttgaaaatc cttgccctta 540  
tcatttgggg tgaaggagac atttctgtcc ttggcttccc acagcccaa cgcagctgtg 600  
gtatgattcc tgggatccaa cgagccctcc tattttcaca gtgttctgat tgctctcaca 660  
gcccaggccc atcgtctgtt ctctgaatgc agccctgttc tcaacaacag ggaggtcatg 720  
gaaccctct gtggaaccca caaggggaga aatgggtgat aaagaatcca gttcctcaaa 780  
accttccctg gcaggtctgg tccctctcct gctgggtggt gctttctctt gcacaccat 840  
cccaccagc ggggagagcc agcaacccaa ccagacagct caggttgtgc atctgatgga 900  
aaccactggg ctcaaacacg tgcctttatc tctgtttat ttttgcgtgt actttgaagc 960  
atggaatctc ttgtttgggg gatcttgggg ctacagtagt gggtaaacaa atgcccaccg 1020  
gccaagaggc cattaacaaa tcgtccttgt cctgaggggc ccagcttgc tcgggcgtgg 1080  
cacagtgggg aatccaaggg tcacagtatg gggagaggtg caccctgcca cctgctaact 1140  
tctcgtaga cacagtgtt ctgcccaggt gacctgttca gcagcagaac aagccagggc 1200  
catggggacg ggggaagttt tcaactggag atggacacca agacaatgaa gatttgttgt 1260  
ccaaataggt caataattct gggagactct tggaaaaaac tgaatatatt caggaccaac 1320  
tctctccctc ccctcatccc acatctcaaa gcagacaatg taaagagaga acatctcaca 1380  
caccagctc gccatgccta ctcatctctg aatttcaggt gccatcactg ctctttcttt 1440  
cttctttgtc atttgagaaa ggatgcagga ggacaattcc cacagataat ctgagggaatg 1500  
cagaaaaacc agggcaggac agttatcgac aatgcattag aacttgggtg gcatcctctg 1560  
tagagggact ccacccctgc tcaacagctt ggcttccagg caagaccaac cacatctggt 1620  
ctctgccttc ggtggccac acacctaagc gtcactgtca ttgccatagc atcatgatgc 1680  
aacacatcta cgtgtagcac tacgacgtta tgtttgggta atgtgggat gaactgcatg 1740  
aggctctgat taaggatgtg gggaggtggg ctgctggtcac tgtcggcctt gcaaggccac 1800  
ctggaggcct gtctgttagc cagtgggtgga ggagcaaggc ttcaggaagg gccagccaca 1860  
tgccatcttc cctgcgatca ggcaaaaaag tggaaataaa aagtcaaac tttatatgca 1920  
tgtgttatgt ccattttgca ggatgaactg agtttaaaag aattttttt tctcttcaag 1980  
ttgctttgtc ttttccatcc tcatcacaag cccttgtttg agtgccttat ccctgagcaa 2040  
tctttcgtat gatggagatg atcattaggt acttttgggt caacctttat tctgttaaat 2100  
atctctgtga aaactaggag aacagagatg agatttgaca aaaaaaatt gaattaaaaa 2160  
taacacagtc tttttaaaac taacatagga aagccttcc tattatttct cttcttagct 2220  
tctccattgt ctaaatcagg aaaacaggaa aacacagctt tctagcagct gcaaaatggt 2280



```

ttaatgcccc ctacatatattt ccatcacctt gaacaatagc tttagcttgg gaatctgaga 2340
tatgatccca gaaaacatct gtctctactt cggctgcaaa acccatgggt taaatctata 2400
tgggtttgtgc attttctcaa ctaaaaatag agatgataat ccgaattctc catatatcca 2460
ctaatacaag acactatttt catactagat tcctgagaca aatactcact gaagggtctg 2520
tttaaaaata aattgtgttt tggctgttct ttgtagataa tgccttctta ttttaggtag 2580
aagctctgga atccctttat tgtgctgttg ctcttatctg caagggtggca agcagttctt 2640
ttcagcagat tttgcccact attcctctga gctgaagtgc ttgcataga tttggcttaa 2700
gcttgaatta gatccctgca aaggcttgct ctgtgatgct agatgtaatt gtaaatgtca 2760
gtaatcactt catgaacgct aaatgagaat gtaagtattt ttaaatgtgt gtatttcaaa 2820
tttgtttgac taattctgga attacaagat ttctatgcag gatttacctt catcctgtgc 2880
atgtttccca aactgtgagg aggggaaggct cagagatcga gcttctcctc tgagttctaa 2940
caaaatgggt ctgtgagggt cagcctttag gaagggtgag cttgtgtgtc ctttgagctt 3000
tctgttatgt gcctatccta ataaactctt aaacacaaaa aaaaaaa 3047

```

&lt;210&gt; 8

&lt;211&gt; 3017

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No.: 1662318CB1

&lt;400&gt; 8

```

cgcaaaactca acccttttcgg aaacaccttc ctcaacaggt tcatgtgtgc ccagctccct 60
aatcaggtcc tggagagcat cagcatcatc gacaccccgg gtatcctgtc gggtgccaag 120
cagagagtga gccgcgcta cgacttcccg gccgtgctgc gctgggtcgc ggagcgcgtg 180
gacctcatca tcctgtctct tgatgcgcac aagctggaga tctcggacga gttctcagag 240
gccatccggc cgttgcgggg ccatgaggac aagatccggc tgggtgctcaa caaggccgac 300
atgggtggga cgcagcagct gatgcgcgtc tacggcgcgc tcatgtgggc gctgggcaag 360
gtgggtggga cgcgcaggt gctgcgcgtc tacatcggct ccttctgtgc ccagcccctc 420
ctgggtgccg acaaccggcg cctcttcgag ctggaggagc aggacctctt ccgcgacatc 480
cagggcctgc cccggcacgc agccttgccg acgtcgaacg acctggtgaa gagggcccgg 540
ctgggtgcgag ttcacgctta catcatcagc tacctgaaga aggagatgcc ctctgtgttt 600
gggaaggaga acaagaagaa gcagctgac ctcaaaactgc ccgtcatctt tgcaagatt 660
cagctggaac atcacatctc ccctggggac tttcctgatt gccagaaaat gcaggagctg 720
ctgatggcgc acgacttcac caagtttcac tcgtgaagc cgaagctgct ggaggcactg 780
gacgagatgc tgacgcacga catcgccaag ctcatgcccc tgctgcggca ggaggagctg 840
gagagcaccc aggtggcgct gcagggggac gcttttgagg gcacccacat gggcccgttt 900
gtggagcggg gacctgacga ggccatggag gacggcgagg agggctcgga cgacgaggcc 960
gagtgggtgg tgaccaagga caagtccaaa tacgacgaga tcttctacaa cctggcgcct 1020
gccgacggca agctgagcgg ctccaaggcc aagacctgga tgggtggggac caagctcccc 1080
aactcagtg cggggcgcat ctggaagctc agcgatgtgg accgcgacgg catgctggat 1140
gatgaagagt tcgcgctggc cagccacctc atcgaggcca agctggaagg ccacgggctg 1200
cccgccaaac tgccccgtcg cctggtgcc aacctcaagc gacgccacaa gggctccgcc 1260
gagtgagcgg ggccccctc ccatggccct gctgtggctc ccagctcca gtcggctgca 1320
cgcacacccc tgctccggct cacacacgcc ctgcctgcc tccctgcca gctgtaagga 1380
ccgggggtct cctcctcac taccgccaga caccocgggt gaagcattta gaggggacca 1440
cgggagggac aaggcttctc tgtccgccct tcaacctcc agcctcacgt tcaattagga 1500
acatcacaca cacttgga cagcaggca tccatccatc cgtcattcat tcaaatattt 1560
attgagcacc tactatgtgc ccagccctgt tctaggcact gggcattacc atagagaaca 1620
aaatagacaa atacatctgc cctcatggaa ggtgacgttc ccaggagagg gcacctacac 1680
agtcacgcaa acacacacta attcctggca gggccccag cccctcccct ggctgagcag 1740
ccctgtggct gaaatgacta gcagataaac agaccccctt ctgctcgcct tcctcctgcc 1800
cagccaggca acacctcaa ccggctccat cacatcctca ggtctcggga ccatgggggg 1860
ctcagagggg agacacacct actgcttctc cagatggggc cctccgcagc cccttcccct 1920
gctcggggaa agcccccaat tctgccaca cccatttatt tcttctcttc cttcttctt 1980
ttcttctctt ccttctctct ttttgtttt gcccccagtt ctgtccacac cccttcccct 2040
tcttctcttc cttctctctt ttttgtttt gccccagtt ctgtccacac cccttcccct 2100
tctgtcctg tcttctctt cttttttgat agaattctgc tctgtcggc aggtggggag 2160
tgacgtggtg agatctcagc tactgcaac ctccacctcc tgggttgaag tgattctcgt 2220
gcctcagcct cctgagtagc tgggactgca ggacgcggc accacgcca gctaattttt 2280

```

```
gtatttgagt agagacgggg tttcaccatg ttggccaggc tggctctgaa ctccgcacatc 2340
caggtgatct gctcgccctcg gcctcccaaa gtgatgggat tacaggcatg agccaccgtg 2400
ccccgcttca caccatttcc tttaaaaagg atcccgtagc aggcagaaaa gcccttcca 2460
tctgtctctc ctgatactgt gcccccttgg agatatattcc gtcctccacc cacgtgtctg 2520
tggctgggaa tgcccagcct gctcctggcc ccctggaagc ctccccacag ctggtaatct 2580
ggacttaagg attgctgggc caccgcctct ctgcctacca ccattccata ttttaagtga 2640
gcccctacgt agaaaggccc cggggcttta ttttagtctc cttttcaggg atgtcgtggg 2700
cgggggaggg ggttcttggg gctacagccc tctccccacc cctaaaggga cgccgacgt 2760
gtttgctgcc ttcaccacat attagtgtt gacctggca ggggacccca tggaaaagat 2820
gggaaagagc aaaatacatg gagacgacgc accctccagg atgctcgtg ggattcccac 2880
gcccaccact gtccccacc ccatggctgg gaggggcctc tgaacggaac agtgtcccca 2940
cagagcgaat aaagcaaggc ttcttcccca aaaaaaaaaa aaaaaaaaaa attggtgcgg 3000
ccgaagttat tcccttc
```

<210> 9  
<211> 1735  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No.: 1996726CB1

```
<400> 9
tcgggaggaa ggagactaca cctgctttgc tgaaatcag gtcgggaagg acgagatgag 60
agtcagagtc aagtggtga cagcgccgc caccatcgg aacaagactt acttggcgg 120
tcaggtgccc tatggagacg tggctactgt agcctgtgag gccaaaggag aacctatgcc 180
caaggtgact tgggtgtccc caaccaacaa ggtgatcccc acctcctctg agaagtatca 240
gatataccaa gatggcactc tccttattca gaaagcccag cgttctgaca gcggcaacta 300
cacctgcttg gtcaggaaca gcgcgggaga ggataggaag acggtgtgga ttcacgtcaa 360
cgtccagcca cccaagatca acggttaacc caaccccatc accaccgtgc gggagatagc 420
agccgggggc agtcggaaac tgattgactg caaagctgaa ggcattccca ccccgagggt 480
gttatgggct ttcccagag gtgtggttct gccagctcca tactatgga accggtcac 540
tgtccatggc aacggttccc tggacatcag gagtttgagg aagagcgact ccgtccagct 600
ggtatgcacg gcacgcaacg agggagggga ggccagggtg atcgtgcagc tcaactgtcct 660
ggagcccacg gagaaacca tcttcacaga cccgatcagc gagaagatca cggccatggc 720
gggccacacc atcagcctca actgctctgc cgcggggacc ccgacacca gctgtgtgtg 780
ggtccttccc atgggcaccg atctgcagag tggacagcag ctgcagcgct tctaccaca 840
ggctgacggc atgtacaca ttagcggctc ctctcgggtg gacgcccggg cctaccgtg 900
cgtggcccg c aatgccgtg gccacacgga gaggtggtc tccctgaagg tgggactgaa 960
gccagaagca aacaagcagt atcataacct ggtcagcatc atcaatggtg agacctgaa 1020
gtcctcctgc acccctccc gggtgggca gggacgttcc tcttgacgc tcccgaatgg 1080
catgcatctg gagggcccc aaacctggg acgcgttctc cttctggaca atggcaccct 1140
cacggttcgt gaggcctcgg tgtttgacag gggtagctat gtatgcagga tggagacgga 1200
atacggccct tcggtacca gcatccccgt gattgtgatc gcctatcctc cccggatcac 1260
cagcgagccc acccgggtca tctacacccg gcccggaac accgtgaaac tgaactgcat 1320
ggctatgggg attcccaaag ctgacatcac gtgggagttc ccggataagt cgcattctgaa 1380
ggcaggggtt caggctcgtc tgtatggaaa cagatttctt cacccccagg gatcactgac 1440
catccagcat gccacacaga gagatgccgg cttctacaag tgcattggaa aaaacattct 1500
cggcagtgac tccaaaacaa cttacatcca cgtcttctga aatgtggatt ccagaatgat 1560
tgcttaggaa ctgacaacaa agcgggggtt gtaagggagg ccaggttggg gaataggagc 1620
tcttaaataa tgtgtcacag tgcattgtgg cctctggtgg gtttcaagtt gaggttgatc 1680
ttgatctaca attgttggga aaagggaagc atgcagacac gagaaggagg gctca 1735
```

<210> 10  
<211> 1016  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature

&lt;223&gt; Incyte ID No.: 2137155CB1

&lt;400&gt; 10

ctgtacggtc	ccctgtggcc	cacgcctagt	gaaaatgata	tcgtacatct	ccctagagat	60
atgggtcacc	tccaggtaga	ttacagagat	aacaggctgc	acccaagtga	agattcttca	120
ctggactcca	ttgcctcagt	tgtggttccc	ataattatat	gcctctctat	tataatagca	180
ttcctattca	tcaatcagaa	gaaacagtgg	ataccactgc	tttgctggta	tcgaacacca	240
actaagcctt	cttccttaaa	taatcagcta	gtatctgtgg	actgcaagaa	aggaaccaga	300
gtccagggtg	acagttccca	gagaatgcta	agaattgcag	aaccagatgc	aagattcagt	360
ggcttctaca	gcatgcaaaa	acagaaccat	ctacaggcag	acaatttcta	ccaaacagtg	420
tgaagaaagg	caactaggat	gaggtttcaa	aagacggaag	acgactaaat	ctgctctaaa	480
aagttaaacta	gaatttgtgc	acttgcttag	tggattgtat	tggattgtga	cttgatgtac	540
agcgctaaga	ccttactggg	atgggctctg	tctacagcaa	tgtgcagaac	aagcattccc	600
acttttctct	aagataactg	accaagtgtt	tottagaacc	aaagttttta	aagttgctaa	660
gatatatttg	ctgtgaagat	agctgtagag	atatttgggg	tggggacagt	gagtttggat	720
ggcgaaatac	accgcacggt	ggtgttggga	agaaaaattt	gtcagcttgg	ctcggggaga	780
aaccttggtg	caactaaagca	gttcagtgtg	ccagagggtta	tttttttccc	attgctctga	840
agactgcact	ggttgcgtga	aactcaggcc	tgaatgagcg	gaaacaaaaa	aagccttgcc	900
ccccgatgcc	ataacacctt	tggaaatccc	agcggccctc	agaaaccttt	tcaggcatcc	960
aggtcttaag	cccaagtatc	tttctatata	gtcccactgc	ggtgagcgtg	ggggag	1016

&lt;210&gt; 11

&lt;211&gt; 2288

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No.: 2268890CB1

&lt;400&gt; 11

caaccagggt	caggctgtgc	tcacagtttc	ctctggcggc	atgtaaaggc	tccacaaagg	60
agttgggagt	tcaaatgagg	ctgctgcgga	cggcctgagg	atggacccca	agccctggac	120
ctgccgagcg	tggcactgag	gcagcggctg	acgctactgt	gagggaaaga	aggttgtgag	180
cagccccgca	ggacccttgg	ccagccctgg	ccccagcctc	tgccggagcc	ctctgtggag	240
gcagagccag	tggagcccg	tgaggcagg	ctgcttggca	gccaccggcc	tgcaactcag	300
gaaccctccc	agaggccatg	gacaggctgc	cccgctgacg	gccagggtga	agcatgtgag	360
gagccgcccc	ggagccaagc	aggagggaag	aggctttcat	agattctatt	cacaaagaat	420
aaccaccatt	ttgcaaggac	catgaggcca	ctgtgcgtga	catgctggtg	gctcggactg	480
ctggctgcc	tgggagctgt	tgcaggccag	gaggacggtt	ttgagggcac	tgaggagggc	540
tcgccaagag	agttcattta	cctaaacagg	tacaagcggg	cgggcgagtc	ccaggacaa	600
tgacactaca	ccttcattgt	gccccagcag	cgggtcacgg	gtgccatctg	cgtcaactcc	660
aaggagcctg	aggtgcttct	ggagaaccga	gtgcataaag	aggagctaga	gctgctcaac	720
aatgagctgc	tcaagcagaa	gcggcagatc	gagacgctgc	agcagctggt	ggaggtggac	780
ggcggcattg	tgagcgaggt	gaagctgctg	cgcaaggaga	gccgcaacat	gaactcgcg	840
gtcacgcagc	tctacatgca	gctcctgcac	gagatcatcc	gcaagcggga	caacgcgttg	900
gagctctccc	agctggagaa	caggatcctg	aaccagacag	ccgacatgct	gcagctggcc	960
agcaagtaca	aggacctgga	gcacaagtac	cagcacctgg	ccacactggc	ccacaaccaa	1020
tcagagatca	tcgcgcagct	tgaggagcac	tgccagaggg	tgccctcggc	caggcccgtc	1080
ccccagccac	cccccgctgc	cccggccccg	gtctaccac	caccaccta	caaccgcata	1140
atcaaccaga	tctctaccaa	cgagatccag	agtgaaccaga	acctgaaggt	gctgccaccc	1200
cctctgcccc	ctatgcccac	tctcaccagc	ctccatctt	ccaccgacaa	gccgtcgggc	1260
ccatggagag	actgcctgca	ggccctggag	gatggccacg	acaccagctc	catctacctg	1320
gtgaagccgg	agaacaccaa	ccgcctcatg	caggtgtggt	gcgaccagag	acacgacccc	1380
gggggctgga	ccgtcatcca	gagacgcctg	gatggctctg	ttacttctt	caggaaactg	1440
gagacgtaca	agcaagggtt	tgggaacatt	gatggcgagt	actggctggg	cctgggagaa	1500
atttactggc	tgacgaacca	aggcaactac	aaactcctgg	tgaccatgga	ggactggacc	1560
ggccgcaaa	tctttgcaga	atacgccagt	ttccgcctgg	aacctgagag	cgagtattat	1620
aagctgcgcc	tggggcgcta	ccatggcaat	gcgggtgact	cctttacatg	gcacaacggc	1680
aagcagttca	ccaccctgga	cagagatcat	gatgtctaca	caaggaaactg	tgccacttac	1740
cagaaggagg	gctggttgga	taacgcctgt	gcccactcca	acctcaacgg	ggtctggtac	1800
cgcggggggc	attaccggag	ccgctaccag	gacggagtct	actgggctga	gttccgagga	1860

ggctcttact cactcaagaa agtgggtgatg atgatccgac cgaaccccaa caccttccac 1920  
taagccagct cccctcctg acctctcgtg gccattgcca ggagccacc ctggtcacgc 1980  
tgccacagc acaagaaca actcctcacc agttcatcct gaggctgga ggaccgggat 2040  
gctggattct gttttccgaa gtcactgcag cggatgatgg aactgaatcg atacggtgtt 2100  
ttctgtccct cttacttttc ttacaccag acagccctc atgtctccag gacaggacag 2160  
gactacagac aactctttct ttaaataaat taagtctcta caataaaaac acaactgcaa 2220  
agtaccttca taatatacat gtgtatgagc ctcccttggt cacgtatgtg tatagcacat 2280  
atatatgg 2288

<210> 12

<211> 3304

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No.: 2305981CB1

<400> 12

ccctcttatg gattcccagc aagcatcagg aaccattgtg caaattgtca tcaataacaa 60  
acacaagcat ggacaagtgt gtgtttccaa tggaaagacc tattctcatg gcgagtcctg 120  
gcacccaaac ctccgggcat ttggcattgt ggagtgtgtg ctatgtactt gtaatgtcac 180  
caagcaagag tgtaagaaaa tccactgccc caatcgatac ccctgcaagt atcctcaaaa 240  
aatagacgga aagtgtctgca aggtgtgtcc aggtaaaaaa gcaaaagaag aacttccagg 300  
ccaaagcttt gacaataaag gctactttct cggggaagaa acgatgcctg tgtatgagtc 360  
tgtattcatg gaggatgggg agacaaccag aaaaatagca ctggagactg agagaccacc 420  
tcaggtagag gtccacgttt ggactattcg aaagggcatt ctccagcact tccatattga 480  
gaagatctcc aagaggatgt ttgaggagct tcctcacttc aagctggtga ccagaacaac 540  
cctgagccag tggaaatctc tcaccgaagg agaagctcag atcagccaga tgtgttcaag 600  
tcgtgtatgc agaacagagc ttgaagattt agtcaagggt ttgtacctgg agagatctga 660  
aaagggccac tgtaggcaa gacagacagt attggatagg gtaaagcaag aaaactcaag 720  
ctgcagctgg actgcaggct tattttgctt aagtcaacag tgccctaaaa ctccaaactc 780  
aatgcagtc aattattcac gccatgcaca gcataatttg ctctttgtg tggagtgggt 840  
tgtcagccct tgaacatctc ctccaaagag actagaagag tcttaaatta tatgtgggag 900  
gaggagggat agaacatcac aacactgctc tagtttcttg gagaatcaca tttctttaca 960  
ggttaaagac aaacaagacc ccagggtttt tatctagaaa gttattcaag tgaaagaaag 1020  
agaaggaat tgcttagtag gagtcttgca gtatagaaca attacttgta tgaaattata 1080  
cctttgaatt ttagaatgtc atgtgttctt ttaaaaaaat tagctcccca tctcctcc 1140  
tactcctc cctcctcct tctctctctc tctctctctc cctctctcac agacacacac 1200  
acacacacac acacacgcac acgcacgtcc aactcacat taaactaaag ctttatttga 1260  
agcaaagcta gccaaaattc tacgttactt ttcccttgac tggatcccaa gtagcttga 1320  
agtttttgtg cccaggagag taaataactg tgaacaagag gctctgccct taggtctctg 1380  
tggctgttta agtcaccaac aatagagtca gggtaaagaa taaaaaact ttcatagcct 1440  
cattcatcca cttagaagtg gtaataattt ttccctaag ataccacttt tcttttcccc 1500  
ctgtacctat gggacttcca gaaagaagtt aaattgagta aaatcatcag aaactgaatc 1560  
catgtaagaa aaaataattg ttgaagaaag aagttgatag aattcaaaaa ggccatcttt 1620  
ttgctttcac atcaataaaa tttaccaagt aatagatcag tactcactaa tttttttgag 1680  
accatagtgt tctggtcaga aaaattatat taaattagta aattctagaa gctctttaa 1740  
agggaggttt tcttcttct ccaattatag gagtgtattt ttactttgca aagtggctcg 1800  
gtcctcatga gcatctgcat gttgactctt cagttaagaa aattgttgtt catttaggga 1860  
ggtggatatt ctgatgaaga tctttatcct aaaccttct actatccttg tcttattcat 1920  
caagcagata ttttagtcaa gaattccaga gaaggctgct cctaaaatgt ctacttgag 1980  
cccaatacca gagcataaac tatccattct ggggtctggc tttagaaatc atctttgtgg 2040  
gaagacctaa ttcttcacag caaggatctc aggcattgct tctagatttg tccctctga 2100  
ggggcaggaa tgaactgtag aaatgtttta aggaccaga aaccccatat gtctcattcc 2160  
atgactatag gtgagagaat tctttcctaa gaggggttga taccaatagg ggaatgtga 2220  
aaatgtttac tctttatgac aacctggcat aaaggagtca attcttatga aagagacaca 2280  
agggctctat ggcagggtt tcttgggaca agactctcac cagcacatca cacacgttct 2340  
ccttgggaaga ggaagcagt acatcccggg tgagagggtc caaagcatta gtttgtgtgt 2400  
gtgtgtgtgt gtgtgtgtgt gtgtgtgtgt gtgtgtgtgt gtggtaaagg ggggaagggtg 2460  
ttatgagggt gctccctccg tcccagaggt ggcagtgatt ccataatgtg gagactagta 2520  
actagatcct aaggcaagaa ggtgtttctc cttctggatg attcatccca aagccttccc 2580

```
accaggtgt tctctgaaag cttagcctta agagaacacg cagagagttt ccctagatat 2640
actcctgcct ccaggtgctg ggacacacct ttgcāaaatg ctgtgggaag caggagctgg 2700
ggagctgtgt taagtcaaag tagaaacct ccagtgtttg gtgttggtga gagaatagga 2760
catagggtaa agaggccaag ctgcctgtag ttagttagaga agaattgatg tggttcttct 2820
tgtgtattta tttgtatcat aaacacttgg aacaacaaag accataagca tcatttagca 2880
gttgtagcca ttttctagtt aactcatgta aacaagtaag agtaacataa cagtattacc 2940
ctttcactgt tctcacagga catgtacctt attatggtac ttatttatgt agtcaactgta 3000
tttctggatt tttaaattaa taaaaaagtt aattttgaāa aaaaaaaaaa aaaaaaaaaa 3060
aaaaaaaaaa aaaaaaaaaa actcgagggg gggcctgtac cgggttcccc gtaacagggt 3120
cgcccttaag attccctggc cgcagttttt ggccgcgttt tggggaacct ctgggtacct 3180
ccttagttgc tcgctaaaat cccctttcgc agcccgttta aaggctgggg ccggccgatt 3240
gccttcccaa tagcctccca tgaatgggaa tggaattgga agggaaattt tggtaaattc 3300
ggta 3304
```

<210> 13  
<211> 708  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No.: 2457612CB1

```
<400> 13
ggaaagccag gaagtgcagg aatcatttca tcaggggccaa taactacacc acccctgagg 60
tcaacaccca ggcctactgg aactcccttg gagagaatag agacagatgt aaagcaacca 120
acagttcctg cctctggaga agaactggaa aatataactg actttagctc aagcccaaca 180
agagaaactg atcctcttgg gaagccaaga ttcaaaggac ctcatgtgcy atacatccaa 240
aagcctgaca acagtccctg ctccattact gactctgtca aacggttccc caaagaggag 300
gccacagagg ggaatgccac cagcccacca cagaaccac ccaccaacct cactgtggtc 360
accgtggaag ggtgcccctt catttgtcat ctgtgactgg gaaaagccac taaatgacac 420
tgtcactgaa tatgaagtta tatccagaga aaatgggtca ttcagtggga agaacaagtc 480
cattcaaatg acaaatcaga cattttccac agtagaaaat ctgaaaccaa acacgagtta 540
tgaattccag gtgaaaccca aaaaccgcgt tgggtgaaggc ccggtcagca acacagtggc 600
attcagtact gaatcagcgg acccagagtg agtgagcagt ttctgcagga gagatgcctc 660
tggactgaag gccgctttgt tcgactcttg ctccaggtgta agggcaac 708
```

<210> 14  
<211> 2040  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No.: 2814981CB1

```
<400> 14
cggccagccg ccgcgcgctg cagctctccg ggacgcccgt gcgccagctg cagaagggcg 60
cctgcccgtt ggggtctccac cagctgagca gccgcgcta caagttcaac ttcattgctg 120
acgtgggtga gaagatcgca ccagccgtgg tccacataga gctcttctctg agacaccgc 180
tgtttggccg caacgtgccc ctgtccagcg gttctggctt catcatgtca gaggccggcc 240
tgatcatcac caatgccac ttgggtgtcca gcaacagtgc tgccccgggc aggcagcagc 300
tcaaggtgca gctacagaat ggggactcct atgaggccac catcaaagac atcgacaaga 360
agtcggacat tgccaccatc aagatccatc ccaagaaaa gctccctgtg ttgttgctgg 420
gtcactcggc cgacctgcgg cctggggagt ttgtggtggc catcggcagt cccttcgccc 480
tacagaacac agtgacaacg ggcacgttca gcactgccca gcgggagggc agggagctgg 540
gctccgggga ctccgacatg gactacatcc agacggatgc catcatcaac tacgggaact 600
ccggggggacc actggtgaac ctggatggcg aggtcattgg catcaaacag ctcaaggtca 660
cggttgcatc ctcttttgcc atcccctcag accgcacac acggttctc acagagttcc 720
aagacaagca gatcaaagac tggaagaagc gcttcacagg catacggatg cggacgatca 780
```

```

caccaagcct ggtggatgag ctgaaggcca gcaacccgga cttcccagag gtcagcagtg 840
gaatttatgt gcaagagggtt gcgccgaatt caccttctca gagaggcggc atccaagatg 900
gtgacatcat cgtcaaggtc aacgggctc ctctagtga ctcgagtga ctgcaggagg 960
ccgtgctgac cgagtctcct ctctactgg aggtgcggcg ggggaacgac gacctcctct 1020
tcagcatcgc acctgagggtg gtcattgtgag gggcgcttc ctccagcgcc aagcgtcaga 1080
gcctgcagac aacggaggggc agcgcccccc cgagatcagg acgaaggacc accgtcggtc 1140
ctcagcaggg cggcagcctc ctctggctg tccggggcag agcggaggct gggcttgccc 1200
aggggcccca atttccgcct ggggagtggt ggatccacat cccggtgccg gggagggaag 1260
cccaacatcc ccttgtacag atgatcctga aagtcacttc caagttctcc ggatattcac 1320
aaaactgcct tccatggagg tccccctcct tcctagcttc ccgcctctgc cctgtgaac 1380
acccatctgc agtatccctt gctcctgccc ctctactgc aggtctggg tgccaagctt 1440
cttccccctt gacaaacgcc cacctgacct gaggccccag cttccctctg ccctaggact 1500
taccaagctg tagggccagg gctgctgcct gccagcctgg ggtccctgga ggacaggatc 1560
catctgatcc ctttggggtg cgggggtggg gtccagccca gacgaggcac tgagtgaatg 1620
ccccctggct gcggagctga gccccgcct gccatgaggt tttcctcccc aggcaggcag 1680
gaggccgctg ggagcacgtg gaaagttggc tgctgcctgg ggaagcttct cctcccaag 1740
gcggccatgg ggcagcctgc agaggacagt ggacgtggag ctgcggggtg tgaggactga 1800
gccggcttcc ccttcccacg cagctctggg atgcagcagc cgctcgcatg gaagtgccgc 1860
ccagaggcat gcaggctgct gggcaccacc cctcatcca gggaacgagt gtgtctcaag 1920
ggcattttgt gacgtttgct gtaaatggat tccagtggt gcttgactg tatgtttctc 1980
tactgtatgg aaaataaagt ttacaagcac aaaaaaaaaa aaaaaaaaaa aaaaaaaagg 2040

```

&lt;210&gt; 15

&lt;211&gt; 2121

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc feature

&lt;223&gt; Incyte ID No.: 3089150CB1

&lt;400&gt; 15

```

gtaaaagctg gttgtgatcg catcatagac tccaaaaaga agtttgataa atgtggtggt 60
tgccgggggaa atggatctac ttgtaaaaaa atatcaggat cagttactag tgcaaacct 120
ggatatcatg atatcatcac aattccaact ggagccacca acatcgaagt gaaacagcgg 180
aaccagaggg gatccaggaa caatggcagc tttcttgcca tcaaagctgc tgatggcaca 240
tatattctta atggtgacta cactttgtcc accttagagc aagacattat gtacaaaggt 300
gttgctcttg ggtacagcgg ctctctcgcg gcattggaag gaattcgag ctttagccct 360
ctcaaagagc ccttgaccat ccaggttctt actgtgggca atgcccttcg acctaaaatt 420
aaatacacct acttcgtaaa gaagaagaag gaatcttca atgctatccc cactttttca 480
gcatgggtca ttgaagagtg gggcgaatgt tctaagtc atgtaattggg ttggcagaga 540
agactggtag aatgccgaga cattaatgga cagcctgctt ccgagtgtgc aaagggaagt 600
aagccagcca gcaccagacc ttgtgcagac catccctgcc cccagtgcca gctgggggag 660
tggtcatcat gttctaagac ctgtgggaag ggttacaaaa aaagaagctt gaagtgtctg 720
tcccatgatg gagggtgtt atctcatgag agctgtgatc ctttaagaa acctaaacat 780
ttcatagact ttgcacaat ggcagaatgc agttaagtgg ttttaagtgg gtttagctttg 840
agggcaaggc aaagtgagga agggctggtg cagggaagc aagaaggctg gagggatcca 900
gcgtatcttg ccagtaacca gtgagtgta tcagtaaggt gggattatgg gggtagatag 960
aaaaggagtt gaatcatcag agtaaaactgc cagttgcaaa tttgatagga tagttagtga 1020
ggattattaa cctctgagca gtgatatagc ataataaagc cccgggcatt attattatta 1080
tttcttttgt tacatctatt acaagtttag aaaaaaaca gcaattgtca aaaaaagtta 1140
gaactattac aaccctgtt tctgtgtact tatcaaatat ttagtatcat gggggttggg 1200
aaatgaaaag taggagaaaa gtgagatttt actaaagcct gttttacttt acctactaa 1260
caatgggggg agaaaggagt acaaatagga tctttgacca gcactgttta tggctgctat 1320
ggtttcagag aatgtttata cattatttct accgagaatt aaaacttcag attgttcaac 1380
atgagagaaa ggctcagcaa cgtgaaataa cgcaaatggc ttcctctttc cttttttgga 1440
ccatctcagt ctttatttgt gtaattcatt ttgaggaaa aacaactcca tgattttatt 1500
caagtgcatt aaagtctaca atggaaaaaa agcagtgaa cattaagatg tgtaaaaagc 1560
tagaggagac acaatgagct tagtacctcc aacttcttt ctttctacc atgtaaccct 1620
gctttgggaa tatggatgta aagaagtaac ttgtgtctca tgaaaatcag tacaatcaca 1680
caaggaggat gaaacgcgg aacaaaaatg aggtgtgtag aacagggtcc cacagggttg 1740
gggacattga gatcacttgt cttgtggtgg ggaggctgct gaggggtagc aggtccatct 1800

```

```

ccagcagctg gtccaacagt cgtatcctgg tgaatgtctg ttcagctctt ctgtgagaat 1860
atgatttttt ccatatgtat atagtaaaat atgttactat aaattacatg tactttataa 1920
gtattgggtt gggtgttcct tccaagaagg actatagtta gtaataaatg cctataataa 1980
catattttat tttatacatt tattttctaag gaaaaaaact tttaaattat atcgcttttg 2040
tggaaagtgc tataaaatag agtatttata caatatatgt tactagaaat aaaagaacac 2100
ttttggaata aaaaaaaaaa a
2121

```

&lt;210&gt; 16

&lt;211&gt; 2900

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc feature

&lt;223&gt; Incyte ID No.: 3206667CB1

&lt;400&gt; 16

```

gaagttttta aaaaaaactac agcagccaaa gaaactatat atatatatat atatatccag 60
aatgattggc tctactgtcc tcattgacct gtttgaacct tagtgacctt cctgtcctc 120
ttcccagttc tctttataga agctctagga gctttcgaaa agccaaagtc tttctgaaga 180
atctgtgctg gacagacata attccctttc tcattgtctc catctttgtt ggtcatggta 240
aggtttttcc atcagcctct gaaaaaatag ttgtgcacaa catctgctca ctggactgtc 300
tgatccaatg taattggctg cgtctggcta attctaagca cttaaagtcta catctaagct 360
atagatttaa gcttgaagct acagattata tcaactatcac caccaccct caccagtg 420
aatcagacag tcagtcact taagttaaag atatttgggt tctttgaatg atttgctgtc 480
acagactatt tggtagaaga aatatttttc acctgagaga ggaagagaaa tttctctagt 540
aacacaaaga gtgagttcta aaaggcatgc ccacatctct ttcgtgcctt aaggatagtg 600
agatgcacac ttatatatat actgtatata tttatatatt tatatatata tttcatatat 660
atatataata ttgcaagctt aagtttgcaa ttcccaaac aatacaaaaa gcaaattaca 720
caccctcacc actgtttctta tctctatagt gatgaaacat taattaggga tcttgctgct 780
tttcttttcc tacacgaagt ttccattaaa gccacagaat aattgatagg gcagctgttt 840
gagaacaggt cccattttca cattagggct ttaaattgaat tagaaactat ttgaggctat 900
aaaaatgtcc ttgagtttgg agcctgagct tccgtgaaat gctgatacat ctgatctatc 960
atgggaattg cagttagaga gagtaaggaa taccattttag tcatctatcc gttcttca 1020
tagcaggaat atgaaagaaa ggcacatgtt taaggaggaat acctaaaggt ttttctaaat 1080
tccaacattt aaaaaggcaat tgtgggctat ttttattttt taatattttg aaataaagtt 1140
tagtgtctag ggctgggagc caggactgat ctccatttc ttttctttg ttcccagcca 1200
tgcttttcta acttgccagg tggacttgac caactacatt accatgctgt gcctcagttt 1260
acctatttgt aaaaaggat taataatact tactacctc acagggtgt tgtgaggctc 1320
tattcatttg ctcttttatt ctctctgtta ttctctgtat gtccagcact ttgtagccat 1380
gggaggaaag ggactataaa agtgtaaat gttaattgaa tgatacggta cctgaaagcc 1440
ttgttttcta gtaagaaaaat gctaccttgc tgtacatact tataaccttg tatttgga 1500
tgagaaatag gtttatattt tcagatctct caaaaatcac atcatttgac caaagaataa 1560
tttaagacac atagaacaga tttttttaat ttatatattc atcctgacca gcttagttct 1620
aataattttt agttgtgagt gattaaaaaa ctttggatca attttggta aacatgccaa 1680
ctttgtagtc tgagtgcacag gcaaggattt ttgggtttaa gatgcacttt tagcacacat 1740
ttgtatttcc cttggcatat cagattgagc taatgggtat gttatttcaa tctaacagcc 1800
accaatctga aattgtattt caaatgttga ttctgtagt ctttaataa taatgaagct 1860
catcttatac attttgcttt caccaattga ttcttcttcc ttttagccca ctattaaaac 1920
atttcttact gaatgggtca tgtaggcttg ctgaacagca cgcattactt gcttcttgaa 1980
gagttcccc attcatccat ttgtcccat agttgctgtg gattatcaag ttttgaagga 2040
actgtacatc ccaacagact gaaacattct aagtgaatg agtataatcc aagtaactgg 2100
tgaacttttg aggtttggag ctggaagaga atggctaaga agatttgaat tataggagg 2160
gaacagaaat catacatgaa aaggttttac tgagaagggtg aaaaccttag atagagggac 2220
atgtgaaaca aaatcatttg aaattttgat tcagacatcc atttccagt gcaaacagca 2280
aagcctgaac ccataaacc aaatgatagg tgaagtggg tggttttatc caatgtctca 2340
agcaagcaat gtctgggaat atcatagagt aacaagtgt ggtcagccaa agaacattc 2400
actgtgtgtg aaccaatacc ataagcatgt attatctaag cacttgatca agaaatatac 2460
atgttgtaca agctctcaat tttgttcatt tattatcaaa tttttaaata acaagtttg 2520
tatgtgattt ggaaaagatg ctttctggat cttaagccag ttgtcagtg aggtcctcag 2580
ggctgcaaat gtcaagacat aaccctgtc ctcacatca tgataccaga tacaggtgaa 2640
tcataggaat ctatctgcct gtgtcctcaa tctcccttca aacaagatgc tgatttgtag 2700

```

```

ggtacttggc aggtttaaatt aaaccagaag aggtgactta ataaaaaagg gaatgacatt 2760
taggggtataa agatctcata agaaatgtaa tatgttaaatt atatcttgct ttatgttgta 2820
aaatatacat tgtttgcgct agaatagaaa tgatttcttt tcaataaaaa gaaagaagga 2880
ctctaaaaaa aaaaaaaaaa

```

```

<210> 17
<211> 2507
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc feature
<223> Incyte ID No.: 3284695CB1

```

```

<400> 17
cagagtgaat cttgtgcctg gtgaccaaag tccctccaaa gtgctcttcc ttctgggtta 60
ttcaagccaa atatctgggt ttccccctct cctcattccc tagcaaacc caattatctt 120
ccaagatagg agatatttcc catccccctc ctttgtaa atctcatctc ccactggaga 180
gcccaggagc ctattcctgg catggatggt ctgtccacac ttgaggctgg gcggtgtatc 240
agacccttca agcagcctgg ctggggccca ggactgagtc tggggtcagc ttccacggtc 300
gcttttccct tcttcaccac ccaccacagc ccaccttgca tgcattggca gccctccac 360
tccagcctga gccatgtgtg cccctgcggg aggaccatt catgccagaa agctggtaac 420
tccctccag catccctgcg gaaggagtca gtttctgaga gtgtgacttt tcaaggcgaa 480
tgatggggaa ggttcccca gtcccacag tggcccacc tctgggacct gcaccagagc 540
ccttctgtgt caccgcgggc tgtgaccca tgcacacac tacgcacaca caactctccg 600
cactgcagta tattcttgcc aaagatttcc tttaaaagca agcactttta ctaattatta 660
ttttgtaaat gtttatcttc ttctgtcttc tccctccctg aatctatctt actgttgttt 720
attgttgaat ctgtgtgtca gccaggagag cgtgtctggt ccttgaaatc gggctgggat 780
gggaaagggt ctgggagaag atgggcaaca aagagccagg gactcatgga catcgagcg 840
acgcagaccc cagcaggttc agtcccgtgc tgcaccagc tgtccagctg ggtgtctgga 900
gggaagaggg cagaggaggg tcatgtccct tcagctgggg gaggggcca gtgagctcca 960
cgtggccttt tcccaaagg agcaagaggg aaggattggg cgagaaaaca atggagaggg 1020
gacctgcgaa ggaaaacagg gaggaagtga gcggtttgat cagcctgcta tcacggtgtt 1080
ctggctctct tatttagcca ggcgcttaag ggacagatac atcacatcct aagtttggga 1140
aaggcctttg acccatgtca tctgagcgtc tctccagta gctctgaaag ctgtggacac 1200
caatggccag gattccttct cccctgggtt ttgaggatcc ctgggtcttc tgagactggc 1260
caggagaggg atgggtgggg cagtgggtgt gtgaaagcag gaggggcagc cctcctggac 1320
aagtgtgatc cccctataaa cggtctcag gaggttagtg agtaggagat tctgccttgt 1380
tctgatgagc ctgtgcagg gctccagggg agcatgctgt ccagggggca cagaagggtg 1440
gtgagtgtga tcaaatctag tctcactccc acttttttag tctcactcct acttttgtcc 1500
accacccctg cctcctggat cttctcccac ttttttttcc agcttttagga cctggggaga 1560
tctgtgagt caaggcagac acccaatcct gcccacacac tcggggctct ccaagaggtt 1620
ggggggcaga gtcccagagc agccctttac ccaggtcca ggccctggaa tcctgagact 1680
cgcttttctc tggccagtgg taacacagga cgtgtgtgcg catgtgcaag tgtggatgta 1740
tgtgtgtgcg tgtgttttgc tcatctcttt agggaaactg ggagtcgggg ttggaggtgc 1800
tgggcaatgg aacttcaaat tcaatgtcgc ccagcagtga gggagtcgg gaggtgaggc 1860
ctgtaggcca accaattggt ggagtctcag cgatagccca ggtgagaagt ggttcaccca 1920
gaggggcagg gtgggggcct cgggcagatc tgtccctctt ggccctctg tcccaaatg 1980
tccaaatgt tggaggacct ctgttcatat cccacgcctg ggctcttgcc agcagtggag 2040
ttactgtaga gggatgtccc aagcttgttt tccaatcagt gttaagctgt ttgaaactct 2100
cctgtgtctg tgttttgttt gtgcgtgtgt gtgagagcac atcagtgtgt gcaggctgtg 2160
tttccccatt tctctcctcc cttcagaccc atcattgaga acaaatgtaa gaaatccctt 2220
cccacaccc tccctgcctc ccaggccctc tgcgggggaa acaagatcac ccagcatcct 2280
tccccacccc agctgtgtat ttatatagat ggaaatatac tttatatatt gtatcatcgt 2340
gcctatagcc gctgccaccg tgtataaate ctggtgtatg ctcttatccc tggacatgaa 2400
tgtattgtac actgacgcgt cccactcct gtacagctgc tttgtttctt tgcaatgcat 2460
tgtatggcct tataaatgat aaagttaaag aaaaaaaaaa aaaaagg 2507

```

```

<210> 18
<211> 2929
<212> DNA

```



&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No.: 3481610CB1

&lt;400&gt; 18

```
aagctcggaa ttcggctcga gatgggttcc tcatcccttc ctgctgcaaa agaagttaac 60
aaaaaacaag tgtgctacaa acacaatttc aatgcaagct cagtttcctg gtgttcaaaa 120
actgttgatg tgtgttgca ctttaccat gctgctaata attcagtcgt gagcccatct 180
atgaagctga atctggttcc tggggaaaac atcacatgcc aggatcccg aataggtgtc 240
ggagagccgg ggaaagtcac ccagaagcta tgccggttct caaacgttcc cagcagccct 300
gagagtccca ttggcgggac catcacttac aaatgtgtag gctcccagtg ggaggagaag 360
agaaatgact gcatctctgc cccaataaac agtctgctcc agatggctaa ggctttgatc 420
aagagccctc ctcaggatga gatgctccct acatacctga aggatcttct tattagcata 480
ggcaaagcgg aacatgaaat cagctcttct cctgggagtc tgggagccat tattaacatc 540
cttgatctgc tctcaacagt tccaacccaa gtaaatcag aaatgatgac gcacgtgctc 600
tctacggtta atatcatcct tggcaagccc gtcttgaaca cctggaaggt tttacaacag 660
caatggacca atcagagttc acagctacta cattcagtg aaagattttc ccaagcatta 720
cagtcaggag atagccctcc attgtccttc tcccaaaacta atgtgcagat gagcagcatg 780
gtaatcaagt ccagccaccc agaaacctat caacagaggt ttgttttccc atactttgac 840
ctctggggca atgtggtcat tgacaagagc tacctagaaa acttgcatgc ggattcgtct 900
attgtcacca tggctttccc aactctccaa gccatccttg ctcaggatat ccaggaaaat 960
aactttgcag agagcttagt gatgacaacc actgtcagcc acaatacgac tatgccattc 1020
aggatttcaa tgacttttaa gaacaatagc ccttcaggcg gcgaaacgaa gtgtgtcttc 1080
tggaacttca ggcttgccaa caacacaggg ggtggggaca gcagtgggtg ctatgttgaa 1140
gaaggtgatg gggacaatgt cacctgtatc tbtgaccacc taacatcatt ctccatcctc 1200
atgtccctg actcccaga tcttagttct ctctgggaa tactcctgga tattatttct 1260
tatgttgggg tgggcttttc catcttgagc ttggcagcct gtctagtgtg ggaagctgtg 1320
gtgtggaaat cgttgaccaa gaatcggact tcttatatgc gccacacctg catagtgaat 1380
atcgctgcct ccttctctgt cgccaacacc tgggtcattg tggctcgtgc catccaggac 1440
aatcgctaca tactctgcaa gacagcctgt gtggctgcca ccttcttcat ccacttcttc 1500
tacctcagcg tcttctctg gatgctgaca ctgggcctca tgctgttcta tcgcctgggt 1560
ttcattctgc atgaaacaag caggtccact cagaaagcca ttgccttctg tcttggctat 1620
ggctgcccac ttggcatctc ggtcatcacg ctgggagcca ccagcccccg ggaagtctat 1680
acgaggaaga atgtctgttg gctcaactgg gaggcacca aggcctgct ggctttcgcc 1740
atcccagcac tgatcattgt ggtggtgaac ataaccatca ctattgtggt catcacciaag 1800
atcctgaggg cttccatttg agacaagcca tgcaagcagg agaagagcag cctgtttcag 1860
atcagcaaga gcattgggg cctcacacca ctcttgggc tcacttgggg ttttggctctc 1920
accactgtgt tcccagggac caacctgtgt ttccatatca tatttgccat cctcaatgtc 1980
ttccagggat tattcatttt actctttgga tgctctgtgg atctgaaggt acaggaagct 2040
ttgtgaata agttttcatt gtcgagatgg tcttcaagc actcaaagtc aacatccctg 2100
ggttcatcca cacctgtgtt ttctatgagt tctccaatca caaggagatt taacaatttg 2160
tttggtaaaa caggaacgta taatgtttcc accccagaag caaccagctc atccctggaa 2220
aactcatcca gtgcttcttc gttgtcaac taagaacagg ataatccaac ctacgtgacc 2280
tcccggggac agtggctgtg cttttaaaaa gagatgcttg caaagcaatg gggaacgtgt 2340
tctcggggca ggtttccggg agcagatgcc aaaaagactt ttcatagag aagaggcttt 2400
cttttgtaaa gacagaataa aaataattgt tatgtttctg ttgttccct cccctcccc 2460
cttgtgtgat accacatgtg tatagtattt aagtgaact caagccctca aggcccaact 2520
tctctgtcta tattgtaata tagaatttcg aagagacatt ttacttttt acacattggg 2580
cacaaagata agctttgatt aaagtagtaa gtaaaaggct acctaggaaa tacttcagtg 2640
aattctaaga aggaaggaag gaagaaagga aggaagaaag ggagggaac agggagaaag 2700
ggaaaaagaa gaaaaagaga tagatgataa taggaacaaa taaagacaaa caacattaag 2760
ggcatatttg taagatttcc atgttaatga tctaataata tcaactcagtg ccacattttg 2820
agaatttttt tttttaatgg gcttcaaaaa ttggaaaact gtgaaagcta agtccattgg 2880
ggggaatgga attacttttg ggggccagta tctttccttt gattgttcc 2929
```

&lt;210&gt; 19

&lt;211&gt; 1725

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No.: 3722004CB1

<400> 19  
gaggcaagaa ttccggcacga gggagagccc gggggcgtgg gggagctcgg ggacctgcgg 60  
accgggggag cccgaacgag ggggatcccg cggcggcgcc agcgaggcgg agggagcaggc 120  
ggtggaggcg aggcaggaag agggagcagga cttggatggg gagaaggggc catcatcggg 180  
agggcctgag gaggaggacg gagaaggctt ctcccttcaaa tacagccccg ggaagctgag 240  
gggaaaccag tacaagaaga tgatgaccaa agaggagctg gaggaggagc agaggattga 300  
gctgacctct gacctcactt ccctgtagca agttccttag gtccctgagcc acaaatattc 360  
ttgcaaatcc ttttgaactg aagaataacg aagttatcct tagcgtcctc cttaaaggctt 420  
ttcctttttg catcttaaaa gcttgagaga taaaacggaa accccagaga ggagtctggg 480  
caggctccca ggggtgatgc tgcctccata aatctgctga gctctagacc ctcaatcagg 540  
acttgtccct tggctagcag gatcctggga acacctttgg ccctgccctg tgtagagatg 600  
ttcatgtctg ttcctgtggg tcactttgtt aagctgaaga gttttaagag gtagagctca 660  
gacctgggac tgggattttt cttaccactc aaacttgcta tccacacacc ctgcacacct 720  
tagataaaaa gaacatttta aaagcagagt tcactttcac tccagtctcc cctcttttgc 780  
cctcactgaa gccaaaccac agaagacttt gaggaatgag agacaaatga ggtagagctc 840  
acctgtgctc accagctcgg tcagggtggg cagccgaccc ctttccctgg gaacccact 900  
tctctctgtg gctggcttgg ttgtcggggg tgagatgcca tattgattac agggcagcaa 960  
agaaccagta ccaggaattt acttgaccat tccccttatt ttcatctag aggaatctcg 1020  
gattcagccc tttcattgct aagacacctt ttactgagg ttcttaccag ctacagccaa 1080  
tctccactct gctatagcag aagcaataat gtttgcttta aaaagatttc ttgacctatg 1140  
cctttttctt gaaagttaga tagattagtt agaacttcag atcatcagat cagtctcaaa 1200  
tgggtttctt ggaattttat atttgacaat atttatacta taccaaactc atttgcagtt 1260  
cttaggtttg ttggttaaaa cattttttta aagcagtaag tttatagaaa atgttttcat 1320  
ttaatggaag gctggggaat gtccagcatc aaccctatg gcattgcatt ccagtggcct 1380  
tctcatctgg gcctggaacc tttggttcag ggcttagggg agaacaggcc acatggcaac 1440  
agccacacag ccttgccctt caacacagag ccacgtgtgc ccaaacagca atagtcatgc 1500  
ccttgctccag gctgggatct aattgataca ataggctcgtt gactccctcc tagtagagct 1560  
atctaggttt gtctggaag tttccgaccc tggcttatag gcaccacacc tcatgtactc 1620  
ctcatggctt ggatctctgt attcagcctt tgttcagtc aataaacttt gagtagatga 1680  
tctcaaaaaa aaaaaaaaaa aggccggcgc aagcttattc ctttt 1725

<210> 20  
<211> 1987  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No.: 3948614CB1

<400> 20  
gacggccagt gcaagctaaa attaaccttc actaaaggga ataagcttgc ggccgcctgg 60  
agctctcggc ctcggcttcg acgacggcaa cttctcgtg ctcatccgg cgggtggagg 120  
gacggacgcg gggctgtaca cctgcaacct gcaccatcac tactgccacc tctacgagag 180  
cctggccgctc cgcttgagg tcaccgacgg cccccggcc accccgcct actgggacgg 240  
cgagaaggag gtgctggcgg tggcgcgcgg cgcacccgcg cttctgacct gcgtgaaccg 300  
cgggcacgtg tggaccgacc ggcacgtgga ggaggctcaa cagggtgtgc actgggaccg 360  
gcagccgccc ggggtcccgc acgaccgcgc ggaccgctg ctggacctct acgctcggg 420  
cgagcgccgc gctacgggg ccctttttct gcgcgaccgc gtggctgtgg gcgcggatgc 480  
ctttgagcgc ggtgacttct cactgcgtat cgagccgctg gaggtcggcg acgagggcac 540  
ctactcctgc cactgcacc accattactg tggcctgcac gaacgcgcg tcttccacct 600  
gacggtcgcc gaacccacg cggagccgcc cccccgggc tctccgggca acggtccag 660  
ccacagcggc gccccaggcc cagacccac actggcgcgc ggccacaacg tcatcaatgt 720  
catcgctccc gagagccgag ccacttctt ccagcagctg ggctacgtgc tggccagct 780  
gctgctcttc atcctgctac tggctcactgt cctcctggcg gcccgaggc gcccgagg 840  
ctacgaatac tcggaccaga agtcgggaaa gtcaaagggg aaggatgta acttggcgga 900  
gttcgctgtg gctgcaggg accagatgct ttacaggagt gaggacatcc agctagatta 960  
caaaaacaac atcctgaagg agaggcgga gctggcccac agccccctgc ctgccaagta 1020

```

catcgaccta gacaaagggt tccggaagga gaactgcaaa tagggaggcc ctgggctcct 1080
ggctgggcca gcagctgcac ctctcctgtc tgtgctcctc ggggcatctc ctgatgctcc 1140
ggggctcacc ccccttccag cggtctgtcc cgctttcctg gaatttggcc tgggcgtatg 1200
cagaggccgc ctccacaccc ctccccagg ggcttggtgg cagcatagcc cccaccctg 1260
cggcctttgc tcacgggtgg cctgcccac cctggcaca accaaaatcc cactgatgcc 1320
catcatgccc tcagaccctt ctgggtctctg cccgctgggg gcctgaagac attcctggag 1380
gacactccca tcagaacctg gcagcccaa aactggggtc agcctcaggg caggagtccc 1440
actcctccag ggctctgtc gtccggggct gggagatgtt cctggaggag gacactccca 1500
tcagaacttg gcagccttga agttggggtc agcctcggca ggagtccac tcctcctggg 1560
gtgctgcctg ccaccaagag ctccccacc tgtaccacca tgtgggactc caggcaccat 1620
ctgttctccc cagggacctg ctgacttgaa tgccagccct tgctcctctg tgttgctttg 1680
ggccacctgg ggctgcaccc cctgcccttt ctctgcccac tccctaccct agccttgctc 1740
tcagccacct tgatagtac tgggtccct gtgacttctg accctgacac ccctcccttg 1800
gactctgcct gggctggagt ctagggtctg ggctacattt ggcttctgta ctggctgagg 1860
acaggggagg gagtgaagtt ggtttgggt ggctgtgtt gccactctca gcacccaca 1920
tttgcattctg ctgggtggacc tgccaccatc acaataaagt ccccatctga tttttaaaaa 1980
aaaaaaa 1987

```

&lt;210&gt; 21

&lt;211&gt; 551

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc feature

&lt;223&gt; Incyte ID No.: 627722CD1

&lt;400&gt; 21

```

Met Glu Glu Ala Glu Leu Val Lys Gly Arg Leu Gln Ala Ile Thr
1 5 10 15
Asp Lys Arg Lys Ile Gln Glu Glu Ile Ser Gln Lys Arg Leu Lys
20 25 30
Ile Glu Glu Asp Lys Leu Lys His Gln His Leu Lys Lys Lys Ala
35 40 45
Leu Arg Glu Lys Trp Leu Leu Asp Gly Ile Ser Ser Gly Lys Glu
50 55 60
Gln Glu Glu Met Lys Lys Gln Asn Gln Gln Asp Gln His Gln Ile
65 70 75
Gln Val Leu Glu Gln Ser Ile Leu Arg Leu Glu Lys Glu Ile Gln
80 85 90
Asp Leu Glu Lys Ala Glu Leu Gln Ile Ser Thr Lys Glu Glu Ala
95 100 105
Ile Leu Lys Lys Leu Lys Ser Ile Glu Arg Thr Thr Glu Asp Ile
110 115 120
Ile Arg Ser Val Lys Val Glu Arg Glu Glu Arg Ala Glu Glu Ser
125 130 135
Ile Glu Asp Ile Tyr Ala Asn Ile Pro Asp Leu Pro Lys Ser Tyr
140 145 150
Ile Pro Ser Arg Leu Arg Lys Glu Ile Asn Glu Glu Lys Glu Asp
155 160 165
Asp Glu Gln Asn Arg Lys Ala Leu Tyr Ala Met Glu Ile Lys Val
170 175 180
Glu Lys Asp Leu Lys Thr Gly Glu Ser Thr Val Leu Ser Ser Ile
185 190 195
Pro Leu Pro Ser Asp Asp Phe Lys Gly Thr Gly Ile Lys Val Tyr
200 205 210
Asp Asp Gly Gln Lys Ser Val Tyr Ala Val Ser Ser Asn His Ser
215 220 225
Ala Ala Tyr Asn Gly Thr Asp Gly Leu Ala Pro Val Glu Val Glu
230 235 240
Glu Leu Leu Arg Gln Ala Ser Glu Arg Asn Ser Lys Ser Pro Thr
245 250 255

```

Glu Tyr His Glu	Pro Val Tyr Ala Asn	Pro Phe Tyr Arg Pro Thr	260	265	270
Thr Pro Gln Arg	Glu Thr Val Thr Pro	Gly Pro Asn Phe Gln Glu	275	280	285
Arg Ile Lys Ile	Lys Thr Asn Gly Leu	Gly Ile Gly Val Asn Glu	290	295	300
Ser Ile His Asn	Met Gly Asn Gly Leu	Ser Glu Glu Arg Gly Asn	305	310	315
Asn Phe Asn His	Ile Ser Pro Ile Pro	Pro Val Pro His Pro Arg	320	325	330
Ser Val Ile Gln	Gln Ala Glu Glu Lys	Leu His Thr Pro Gln Lys	335	340	345
Arg Leu Met Thr	Pro Trp Glu Glu Ser	Asn Val Met Gln Asp Lys	350	355	360
Asp Ala Pro Ser	Pro Lys Pro Arg Leu	Ser Pro Arg Glu Thr Ile	365	370	375
Phe Gly Lys Ser	Glu His Gln Asn Ser	Ser Pro Thr Cys Gln Glu	380	385	390
Asp Glu Glu Asp	Val Arg Tyr Asn Ile	Val His Ser Leu Pro Pro	395	400	405
Asp Ile Asn Asp	Thr Glu Pro Val Thr	Met Ile Phe Met Gly Tyr	410	415	420
Gln Gln Ala Glu	Asp Ser Glu Glu Asp	Lys Lys Phe Leu Thr Gly	425	430	435
Tyr Asp Gly Ile	Ile His Ala Glu Leu	Val Val Ile Asp Asp Glu	440	445	450
Glu Glu Glu Asp	Glu Gly Glu Ala Glu	Lys Pro Ser Tyr His Pro	455	460	465
Ile Ala Pro His	Ser Gln Val Tyr Gln	Pro Ala Lys Pro Thr Pro	470	475	480
Leu Pro Arg Lys	Arg Ser Glu Ala Ser	Pro His Glu Asn Thr Asn	485	490	495
His Lys Ser Pro	His Lys Asn Ser Ile	Ser Leu Lys Glu Gln Glu	500	505	510
Glu Ser Leu Gly	Ser Pro Val His His	Ser Pro Phe Asp Ala Gln	515	520	525
Thr Thr Gly Asp	Gly Thr Glu Asp Pro	Ser Leu Thr Ala Leu Arg	530	535	540
Met Arg Met Ala	Lys Leu Gly Lys Lys	Val Ile	545	550	

&lt;210&gt; 22

&lt;211&gt; 99

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No.: 1556751CD1

&lt;400&gt; 22

Met Glu Ala Leu Ala	Asn Val Asn Phe	Pro Arg Lys Ser Phe Arg	1	5	10	15
Pro Glu Asp Ala Gly	Lys Glu Ser Gly	Ser Gln Gly Gly Phe Cys	20	25	30	35
Val Pro Ala Ala Arg	Pro Gln Thr Met	Val Thr Gly Pro Ser Cys	35	40	45	50
Ser Ser Pro Gly Leu	Gln Asn Phe Ser	Pro Gln Arg Lys Glu Asn	50	55	60	65
Arg Ala Cys Ala Cys	Trp Gln Asn Ala	Gly Pro Ala Pro Lys Asn	65	70	75	80
Pro Met Cys Val Arg	Leu Lys Val Gly	Arg Pro Gln Ala Ser Gln				

Arg Lys Leu Lys Glu Thr Gly Leu Cys  
80  
95

85

90

<210> 23  
<211> 493  
<212> PRT  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No.: 2268890CD1

<400> 23  
Met Arg Pro Leu Cys Val Thr Cys Trp Trp Leu Gly Leu Leu Ala  
1 5 10 15  
Ala Met Gly Ala Val Ala Gly Gln Glu Asp Gly Phe Glu Gly Thr  
20 25 30  
Glu Glu Gly Ser Pro Arg Glu Phe Ile Tyr Leu Asn Arg Tyr Lys  
35 40 45  
Arg Ala Gly Glu Ser Gln Asp Lys Cys Thr Tyr Thr Phe Ile Val  
50 55 60  
Pro Gln Gln Arg Val Thr Gly Ala Ile Cys Val Asn Ser Lys Glu  
65 70 75  
Pro Glu Val Leu Leu Glu Asn Arg Val His Lys Gln Glu Leu Glu  
80 85 90  
Leu Leu Asn Asn Glu Leu Leu Lys Gln Lys Arg Gln Ile Glu Thr  
95 100 105  
Leu Gln Gln Leu Val Glu Val Asp Gly Gly Ile Val Ser Glu Val  
110 115 120  
Lys Leu Leu Arg Lys Glu Ser Arg Asn Met Asn Ser Arg Val Thr  
125 130 135  
Gln Leu Tyr Met Gln Leu Leu His Glu Ile Ile Arg Lys Arg Asp  
140 145 150  
Asn Ala Leu Glu Leu Ser Gln Leu Glu Asn Arg Ile Leu Asn Gln  
155 160 165  
Thr Ala Asp Met Leu Gln Leu Ala Ser Lys Tyr Lys Asp Leu Glu  
170 175 180  
His Lys Tyr Gln His Leu Ala Thr Leu Ala His Asn Gln Ser Glu  
185 190 195  
Ile Ile Ala Gln Leu Glu Glu His Cys Gln Arg Val Pro Ser Ala  
200 205 210  
Arg Pro Val Pro Gln Pro Pro Pro Ala Ala Pro Pro Arg Val Tyr  
215 220 225  
Gln Pro Pro Thr Tyr Asn Arg Ile Ile Asn Gln Ile Ser Thr Asn  
230 235 240  
Glu Ile Gln Ser Asp Gln Asn Leu Lys Val Leu Pro Pro Pro Leu  
245 250 255  
Pro Thr Met Pro Thr Leu Thr Ser Leu Pro Ser Ser Thr Asp Lys  
260 265 270  
Pro Ser Gly Pro Trp Arg Asp Cys Leu Gln Ala Leu Glu Asp Gly  
275 280 285  
His Asp Thr Ser Ser Ile Tyr Leu Val Lys Pro Glu Asn Thr Asn  
290 295 300  
Arg Leu Met Gln Val Trp Cys Asp Gln Arg His Asp Pro Gly Gly  
305 310 315  
Trp Thr Val Ile Gln Arg Arg Leu Asp Gly Ser Val Asn Phe Phe  
320 325 330  
Arg Asn Trp Glu Thr Tyr Lys Gln Gly Phe Gly Asn Ile Asp Gly  
335 340 345  
Glu Tyr Trp Leu Gly Leu Glu Asn Ile Tyr Trp Leu Thr Asn Gln  
350 355 360

Gly	Asn	Tyr	Lys	Leu	Leu	Val	Thr	Met	Glu	Asp	Trp	Ser	Gly	Arg	
				365					370					375	
Lys	Val	Phe	Ala	Glu	Tyr	Ala	Ser	Phe	Arg	Leu	Glu	Pro	Glu	Ser	
				380					385					390	
Glu	Tyr	Tyr	Lys	Leu	Arg	Leu	Gly	Arg	Tyr	His	Gly	Asn	Ala	Gly	
				395					400					405	
Asp	Ser	Phe	Thr	Trp	His	Asn	Gly	Lys	Gln	Phe	Thr	Thr	Leu	Asp	
				410					415					420	
Arg	Asp	His	Asp	Val	Tyr	Thr	Gly	Asn	Cys	Ala	His	Tyr	Gln	Lys	
				425					430					435	
Gly	Gly	Trp	Trp	Tyr	Asn	Ala	Cys	Ala	His	Ser	Asn	Leu	Asn	Gly	
				440					445					450	
Val	Trp	Tyr	Arg	Gly	Gly	His	Tyr	Arg	Ser	Arg	Tyr	Gln	Asp	Gly	
				455					460					465	
Val	Tyr	Trp	Ala	Glu	Phe	Arg	Gly	Gly	Ser	Tyr	Ser	Leu	Lys	Lys	
				470					475					480	
Val	Val	Met	Met	Ile	Arg	Pro	Asn	Pro	Asn	Thr	Phe	His			
				485					490						